

AMC PAMPHLET
AMC PAMPHLET
AMC PAMPHLET
AMC PAMPHLET

DARCOM-P 706-136
DARCOM-P 706-137
DARCOM-P 706-138
DARCOM-P 706-139

ENGINEERING DESIGN HANDBOOK

SERVOMECHANISMS

SECTION 1, THEORY

SECTION 2, MEASUREMENT AND SIGNAL CONVERTERS

SECTION 3, AMPLIFICATION

SECTION 4, POWER ELEMENTS AND SYSTEM DESIGN



THIS IS A REPRINT WITHOUT CHANGE OF ORDP 20-136, REDESIGNATED AMCP 706-136

ENGINEERING DESIGN HANDBOOK

SERVOMECHANISMS

SECTION 1, THEORY



HEADQUARTERS
UNITED STATES ARMY MATERIEL COMMAND
WASHINGTON, D. C. 20315

30 April 1965

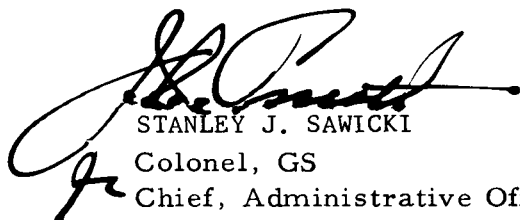
AMCP 706-136, Servomechanisms, Section 1, Theory, forming part of the Army Materiel Command Engineering Design Handbook Series, is published for the information and guidance of all concerned.

(AMCRD)

FOR THE COMMANDER:

SELWYN D. SMITH, JR.
Major General, USA
Chief of Staff

OFFICIAL:



STANLEY J. SAWICKI
Colonel, GS
Chief, Administrative Office

DISTRIBUTION: Special

LIST OF ILLUSTRATIONS

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 1-1 | Elements of a feedback control system | 1-1 |
| 1-2 | Elements of an open-loop control system | 1-2 |
| 1-3 | General diagram of a feedback control system | 1-4 |
| 1-4 | Unity-feedback system main loop | 1-4 |
| 1-5 | Basic unity-feedback system | 1-4 |
| 2-1 | Speed control of shunt d-c motor | 2-3 |
| 2-2 | Transient input functions | 2-6 |
| 2-3 | Transient responses for system $\frac{d^3x}{dt^3} + 2 \frac{d^2x}{dt^2} + 2 \frac{dx}{dt} + x = \frac{d^2y}{dt^2}$ | 2-6 |
| 2-4 | Straight-line crossing course | 2-8 |
| 3-1 | Exponential functions e^{-s} and $1 - e^{-s}$ | 3-4 |
| 3-2 | Cubic chart | 3-6 |
| 3-3 | Cubic chart | 3-7 |
| 3-4 | Quartic chart | 3-8 |
| 3-5 | Sketch of the quartic chart | 3-11 |
| 3-6 | Block diagram manipulation and reduction "rules" | 3-19 |
| 3-7 | Block diagram examples | 3-22 |
| 3-8 | Signal-flow graph in three variables | 3-25 |
| 3-9 | Signal-flow graph of order one | 3-26 |
| 3-10 | Signal-flow graph of order two | 3-26 |
| 3-11 | Signal-flow graph showing addition of parallel branches | 3-26 |
| 3-12 | Signal-flow graph showing multiplication of cascaded branches | 3-27 |
| 3-13 | Signal-flow graph showing termination shifted one node forward | 3-27 |
| 3-14 | Signal-flow graph showing origin shifted one node back- ward | 3-27 |
| 3-15 | Signal-flow graph showing elimination of a self-loop | 3-28 |
| 3-16 | Signal-flow graph showing reduction of second-order graph | 3-28 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 3-17 | Geometry of general trapezoid for approximating $Re [F(j\omega)]$ | 3-29 |
| 3-18 | Real-part function for $W(s) = 1.4s + 0.14/(s^3 + s^2 + 1.4s + 0.14)$ | 3-31 |
| 3-19 | Trapezoidal approximation for $Re[W(j\omega)]$ | 3-32 |
| 3-20 | Impulse response from Floyd's method $W(s) = 1.4s + 0.14/(s^3 + s^2 + 1.4s + 0.14)$ | 3-32 |
| 3-21 | Gain-phase loci of constant real part of $W = G/(1 + G)$ | 3-33 |
| 3-22 | Elements of analog computers | 3-40 |
| 3-23 | Computer diagram for system of Fig. 3-7A | 3-41 |
| 4-1 | Single-loop system — block diagram | 4-1 |
| 4-2 | Locus of s for the Nyquist criterion | 4-5 |
| 4-3 | Locus of $[\lambda(\lambda + 1)^2]^{-1}$ for $\lambda = j\mu$ | 4-5 |
| 4-4 | Distortion of locus of $[\lambda(\lambda + 1)^2]^{-1}$ | 4-5 |
| 4-5 | Locus of $\frac{1}{s} \left(\frac{1+s}{1-s} \right)$ | 4-6 |
| 4-6 | Distortion of locus of $\frac{1}{s} \left(\frac{1+s}{1-s} \right)$ | 4-6 |
| 4-7 | Root-loci plots | 4-10 |
| 4-8 | Poles for $G(s) = 300K/[s(s + 10)(s + 30)]$ | 4-15 |
| 4-9 | Asymptotes and real axis behavior for $G(s) = 300K/[s(s + 10)(s + 30)]$ | 4-15 |
| 4-10 | Construction for determining imaginary axis crossing $G(s) = 300K/[s(s + 10)(s + 30)]$ when $0 < K < \infty$ | 4-15 |
| 4-11 | Complete root locus for $G(s) = 300K/[s(s + 10)(s + 30)]$ when $0 < K < \infty$ | 4-15 |
| 5-1 | Single-loop unity-feedback system | 5-2 |
| 5-2 | Bandwidth measures from magnitude of closed-loop frequency response $C(j\omega)/R(j\omega)$ | 5-2 |
| 5-3 | Bandwidth measure from phase of closed-loop frequency response $C(j\omega)/R(j\omega)$ | 5-2 |
| 5-4 | Bandwidth measure from magnitude of error-to-input frequency response $E(j\omega)/R(j\omega)$ | 5-3 |
| 5-5 | Bandwidth measures from open-loop frequency response | 5-3 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 5-6 | Direct and inverse polar plots of $G(j\omega) = \{j\omega[(j\omega)^2 + 0.6j\omega + 1]\}^{-1}$ | 5-4 |
| 5-7 | Asymptotes and true magnitude curves for the first-order factor $(Tj\omega + 1)^{\pm 1}$ | 5-5 |
| 5-8 | Asymptotes and true curves for the second-order factor $[(j\frac{\omega}{\omega_n})^2 + 2\zeta j\frac{\omega}{\omega_n} + 1]^{-1}$ | 5-6 |
| 5-9 | Phase-angle curves for the first-order factor $(Tj\omega + 1)^{\pm 1}$ | 5-7 |
| 5-10 | Phase-angle curves for the second-order factor $[(j\frac{\omega}{\omega_n})^2 + 2\zeta j\frac{\omega}{\omega_n} + 1]$ | 5-8 |
| 5-11 | Magnitude plots for $G(j\omega) = K \frac{[0.2j\omega + 1]}{j\omega \left[\left(j\frac{\omega}{10} \right)^2 + 0.6j\frac{\omega}{10} + 1 \right]}$ | 5-9 |
| 5-12 | Angle plots for $G(j\omega) = K \frac{(0.2j\omega + 1)}{j\omega \left[\left(j\frac{\omega}{10} \right)^2 + 0.6j\frac{\omega}{10} + 1 \right]}$ | 5-9 |
| 5-13 | Gain-phase plot of $G(j\omega) = 6.5 \frac{(0.2j\omega + 1)}{(j\omega) \left[\left(j\frac{\omega}{10} \right)^2 + 0.6j\frac{\omega}{10} + 1 \right]}$ | 5-10 |
| 5-14 | Closed-loop response construction on the G plane | 5-11 |
| 5-15 | Closed-loop response construction on the G^{-1} plane | 5-11 |
| 5-16 | Contours of constant M in the G plane | 5-13 |
| 5-17 | Contours of constant phase in the G plane | 5-13 |
| 5-18 | Contours of constant M and constant ϕ in the G^{-1} plane | 5-13 |
| 5-19 | Chart showing symmetry of M - N contours about phase of 180 degrees (Nichols Chart) | 5-13 |
| 5-20 | Nichols chart | 5-14 |
| 5-21 | Nonunity-feedback system | 5-15 |
| 5-22 | System equivalent to the system of Fig. 5-21 | 5-15 |
| 5-23 | Construction for gain determination on direct (G/K) plane | 5-16 |
| 5-24 | Construction for gain determination on inverse (KG^{-1}) plane | 5-16 |
| 5-25 | Direct-plane determination of K for $M_p = 1.6$, $G(j\omega) = K \{j\omega[(j\omega)^2 + 0.6j\omega + 1]\}^{-1}$ | 5-17 |
| 5-26 | Inverse-plane determination of K for $M_p = 1.6$, $G(j\omega) = K \{j\omega[(j\omega)^2 + 0.6j\omega + 1]\}^{-1}$ | 5-18 |
| 5-27 | Construction for gain determination on gain-phase plane | 5-19 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 5-28 | Gain-phase plane determination of K for $M_p = 1.5$, $G(j\omega)$ $= K \frac{(0.2j\omega + 1)}{j\omega \left[\left(j \frac{\omega}{10} \right)^2 + 0.6 j \frac{\omega}{10} + 1 \right]}$ | 5-19 |
| 5-29 | Phase margin and gain margin | 5-20 |
| 5-30 | Approximate closed-loop magnitude response of unity-feed-back system, $G(j\omega) = K[j\omega(j\omega + 1)]^{-1}$ | 5-22 |
| 5-31 | Loci of characteristic parameters of second-order root | 5-24 |
| 5-32 | Gain determination from root locus, $G(s) = K(0.2s + 1)/[s(s + 1)(0.1s + 1)]$ | 5-25 |
| 6-1 | Compensation in a single loop | 6-1 |
| 6-2 | Change in open-loop response produced by compensation illustrating downward motion of M_p , contour for gain increase | 6-2 |
| 6-3 | Universal lag functions | 6-2 |
| 6-4 | Lag-compensation procedure | 6-4 |
| 6-5 | Universal lead functions | 6-4 |
| 6-6 | Lead-compensation procedure | 6-5 |
| 6-7 | Magnitude curves for lag-compensation procedure employing phase margin | 6-8 |
| 6-8 | Phase curves for lag-compensation procedure employing phase margin | 6-8 |
| 6-9 | Magnitude curves for lead-compensation procedure employing phase margin | 6-9 |
| 6-10 | Phase curves for lead-compensation procedure employing phase margin | 6-9 |
| 6-11 | General feedback-compensation configuration | 6-11 |
| 6-12 | Cascade equivalent of feedback-compensation configuration | 6-11 |
| 6-13 | Open-minor-loop asymptote for feedback compensation procedure | 6-13 |
| 6-14 | Closed-minor-loop magnitude for feedback compensation procedure | 6-14 |
| 6-15 | Closed-minor-loop phase angle for feedback compensation procedure | 6-14 |
| 6-16 | D-C compensation networks | 6-17 |
| 6-17 | Equivalent circuit elements for carrier-frequency networks | 6-18 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 6-18 | Parallel-T notch filter | 6-19 |
| 6-19 | Inertia damper | 6-20 |
| 6-20 | Block diagram of inertia damper (motor damping negligible) | 6-20 |
| 6-21 | Hydro-mechanical compensation network | 6-20 |
| 6-22 | Schematic diagram of a pneumatic controller | 6-21 |
| 6-23 | Controller block diagram | 6-21 |
| 6-24 | Schematic diagram of a proportional controller | 6-22 |
| 6-25 | Schematic diagram of a proportional plus integral controller | 6-22 |
| 6-26 | Schematic diagram of proportional plus derivative controller | 6-23 |
| 7-1 | Normalized curves yielding time for 10-percent transient response corresponding to combinations of various time constants | 7-2 |
| 7-2 | Normalized curves yielding time for 40-percent transient response corresponding to combinations of various time constants | 7-2 |
| 7-3 | Normalized curves yielding time for 80-percent transient response corresponding to combinations of various time constants | 7-3 |
| 7-4 | Normalized curves yielding time-interval ratios of the transient response corresponding to combinations of various time constants | 7-3 |
| 7-5 | Normalized curves yielding the time interval between 10- and 80-percent response of the transient corresponding to combinations of various time constants | 7-4 |
| 7-6 | Rectangular approximation to step response | 7-5 |
| 7-7 | Rectangular approximation to impulse response | 7-5 |
| 7-8 | Triangular approximation to time function | 7-5 |
| 7-9 | Triangle function | 7-6 |
| 7-10 | Partial derivatives for Linvill's procedure | 7-7 |
| 7-11 | Dimensionless transient error-response curves of a second-order servomechanism to a unit-ramp input | 7-17 |
| 7-12 | Transient error-response curves of a second-order servomechanism to a unit-step input | 7-18 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 7-13 | Transient output-response curves of a second-order servomechanism to a unit-step input | 7-18 |
| 7-14 | Peak magnitude M_p versus damping ratio ζ for a second-order servomechanism | 7-18 |
| 7-15 | Overshoot variation with ζ | 7-19 |
| 7-16 | Typical open-loop asymptote function | 7-21 |
| 7-17 | Sketches showing nomenclature used to describe various characteristics of servomechanism performance | 7-22 |
| 7-18 | Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c | 7-24 |
| 7-19 | Sketch of open-loop asymptote function | 7-42 |
| 8-1 | Configuration for ISE minimization | 8-3 |
| 8-2 | Step-function responses of the optimum unit-numerator transfer systems, second to eighth orders | 8-4 |
| 8-3 | Frequency responses of the optimum unit-numerator transfer systems | 8-5 |
| 8-4 | Step-function responses of the optimum zero-velocity-error systems, second to sixth orders | 8-6 |
| 8-5 | Configuration for MSE minimization | 8-6 |
| 8-6 | Configuration for MSE minimization | 8-8 |
| 9-1 | Sampled-data system | 9-1 |
| 9-2 | Model of sampled-data system | 9-2 |
| 9-3 | Train of unit impulses which represents the carrier $\Delta(t)$ | 9-3 |
| 9-4 | Action of sampled-clamper | 9-3 |
| 9-5 | Simplified picture of sampled-data system | 9-3 |
| 9-6 | Operation of sampling switch | 9-3 |
| 9-7 | Relations between s plane and z plane | 9-5 |
| 9-8 | Sampling a smoothed sampled signal | 9-6 |
| 9-9 | Sampling a filtered continuous signal | 9-6 |
| 9-10 | A sampled-data feedback system | 9-6 |
| 9-11 | Relations between s and $G^*(s)$ for application of Nyquist criterion | 9-7 |
| 9-12 | Comparison between discrete and continuous compensation | 9-9 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 9-13 | Sampled-time function | 9-11 |
| 9-14 | Step response of sampled-data system | 9-11 |
| 9-15 | Determination of $c(t)$ between sampling instants by sampling at $n\Omega$ rad/sec | 9-12 |
| 9-16 | Determination of $c(t)$ between sampling instants through the use of an artificial delay | 9-12 |
| 10-1 | Nonlinear feedback control system | 10-2 |
| 10-2 | Dimensionless representation of contactor characteristics | 10-2 |
| 10-3 | Plot of the describing function N | 10-3 |
| 10-4 | Nonlinear characteristics with dead zone and saturation | 10-4 |
| 10-5 | Describing function for dead zone | 10-4 |
| 10-6 | Describing function for saturation | 10-4 |
| 10-7 | Describing function for saturation and dead zone | 10-4 |
| 10-8 | Hysteresis nonlinearity | 10-5 |
| 10-9 | Describing function for hysteresis-type nonlinear element | 10-5 |
| 10-10 | Simplified nonlinear system | 10-5 |
| 10-11 | Stability determination with describing function | 10-5 |
| 10-12 | Contactor servomechanism study | 10-7 |
| 10-13 | Contactor servomechanism | 10-7 |
| 10-14 | Degree of stability variation with input amplitude for contactor servomechanism | 10-7 |
| 10-15 | Phase portrait of linear second-order system with $\zeta = 0.5$ | 10-9 |
| 10-16 | Portrait in the vicinity of a stable node | 10-10 |
| 10-17 | Portrait in the vicinity of an unstable node | 10-10 |
| 10-18 | Portrait in the vicinity of a stable focus | 10-10 |
| 10-19 | Portrait in the vicinity of an unstable focus | 10-11 |
| 10-20 | Portrait in the vicinity of a center | 10-11 |
| 10-21 | Portrait in the neighborhood of a saddle point | 10-11 |
| 10-22 | Types of singularities | 10-12 |
| 10-23 | Portrait with soft self-excitation | 10-12 |
| 10-24 | Portrait with hard self-excitation | 10-12 |

LIST OF ILLUSTRATIONS

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 11-1 | Electrical representation of a potentiometer | 11-1 |
| 11-2 | Wire-wound element and slider | 11-2 |
| 11-3 | Potentiometer linearity performance compared with ideal performance | 11-5 |
| 11-4 | Potentiometer linearity performance compared with the best straight line | 11-5 |
| 11-5 | Potentiometer with a typical trimming circuit | 11-6 |
| 11-6 | Effect of loading on potentiometer output | 11-6 |
| 11-7 | Load compensation by the addition of a series resistor | 11-7 |
| 11-8 | Load compensation by the addition of a potentiometer tap and a parallel resistor | 11-7 |
| 11-9 | Straight-line approximation of smooth nonlinear function — no change in slope sign | 11-9 |
| 11-10 | Resistance and shaft angle increments for straight-line approximations to nonlinear function — 3 regions of unchanging slope sign | 11-9 |
| 11-11 | Resistance diagram and resultant circuit diagram for the tapped nonlinear function of Fig. 11-10 | 11-10 |
| 11-12 | Equivalent resistance diagram showing a resistance branch for each region of unchanging slope sign | 11-10 |
| 11-13 | Tapped potentiometer for generating a nonlinear function .. | 11-11 |
| 11-14 | Circuit for measuring equivalent noise resistance, ENR | 11-13 |
| 11-15 | Wattage derating curve for rheostat-connected metal-base potentiometers | 11-14 |
| 11-16 | Wattage derating curve for rheostat-connected bakelite-base potentiometers | 11-14 |
| 11-17 | Electrical characteristics of a typical synchro | 11-17 |
| 11-18 | Electrical connections for transmitter-receiver system | 11-20 |
| 11-19 | Electrical connections for transmitter-transformer system .. | 11-21 |
| 11-20 | Electrical connections for a differential transmitter system .. | 11-23 |
| 11-21 | Electrical connections for a differential receiver system | 11-24 |
| 11-22 | Connections for zeroing a transmitter or receiver | 11-26 |
| 11-23 | Connections for zeroing a control transformer | 11-27 |
| 11-24 | Connections for zeroing a differential synchro unit | 11-28 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 11-25 | Torque and rotor current as a function of displacement angle | 11-28 |
| 11-26 | Sources of error in a synchro unit | 11-29 |
| 11-27 | Induction resolver windings | 11-37 |
| 11-28 | Resolver arrangement for vector addition | 11-38 |
| 11-29 | Connections for booster amplifier to primary winding of resolver | 11-39 |
| 11-30 | Induction potentiometer | 11-41 |
| 11-31 | Magnetomotive force and flux relationships of a toroid-wound rotary transformer | 11-42 |
| 11-32 | Voltage distribution in a toroid-wound rotary transformer | 11-43 |
| 11-33 | Toroid-wound rotary transformers in back-to-back circuit | 11-44 |
| 11-34 | Winding schematics of typical microsins | 11-45 |
| 11-35 | Linear variable differential transformer | 11-48 |
| 11-36 | Approximate equivalent circuit of linear differential transformer | 11-49 |
| 11-37 | Linear variable differential transformers used in remote control | 11-50 |
| 11-38 | Construction of drag-cup a-c tachometer generator | 11-52 |
| 11-39 | Schematic diagram of drag-cup a-c tachometer generator | 11-53 |
| 11-40 | Equivalent circuits of drag-cup a-c tachometer generator | 11-53 |
| 11-41 | Pictorial diagrams of the three basic types of gyro units | 11-62 |
| 11-42 | Plot for determining the value of the earth-rate component along the local vertical at any latitude | 11-67 |
| 11-43 | Single-axis feedback system using an integrating gyro unit to drive the gyro supporting member according to an angular velocity command signal | 11-72 |
| 11-44 | Feedback system using an integrating gyro unit to measure the angular velocity about the gyro unit input axis | 11-74 |
| 11-45 | Determination of present position of an aircraft by means of inertial guidance | 11-77 |
| 11-46 | Physical arrangement employed in a typical vertical indicating system | 11-79 |
| 11-47 | Functional diagram, showing single-axis operation only, of a vertical indicating system | 11-80 |
| 11-48 | Binary coding disc | 11-84 |
| 11-49 | Cyclic coding disc | 11-86 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 11-50 | Voltage coder operating on time-interval principle | 11-87 |
| 11-51 | Voltage coder operating on voltage-comparison principle | 11-88 |
| 11-52 | Simple differential linkage | 11-89 |
| 11-53 | Pressure transmitter | 11-90 |
| | | |
| 12-1 | Modulator waveforms | 12-1 |
| 12-2 | Chopper elements | 12-2 |
| 12-3 | Chopper modulator | 12-2 |
| 12-4 | Relationship between driving voltage and chopper output | 12-3 |
| 12-5 | Half-wave connection | 12-5 |
| 12-6 | Half-wave connection | 12-5 |
| 12-7 | Push-pull half-wave connection | 12-5 |
| 12-8 | Differential connection | 12-5 |
| 12-9 | Basic circuit of second-harmonic modulator | 12-6 |
| 12-10 | Operation of cores for positive control signal | 12-6 |
| 12-11 | Secondary voltage for $i_c = I_c$ condition | 12-7 |
| 12-12 | Control characteristics of commercial magnetic modulator | 12-8 |
| 12-13 | Diode modulator | 12-8 |
| 12-14 | Diode modulator .. | 12-9 |
| 12-15 | Triode modulator | 12-9 |
| 12-16 | Bridge modulator | 12-9 |
| 12-17 | Modulator rms output as a function of amplitude of V for fixed amplitude of e | 12-10 |
| 12-18 | Diode demodulator | 12-11 |
| 12-19 | Full-wave diode demodulator | 12-13 |
| 12-20 | Full-wave (ring) diode demodulator | 12-13 |
| 12-21 | Triode demodulator | 12-13 |
| 12-22 | Keyed demodulator | 12-14 |
| 12-23 | Full-wave chopper demodulator | 12-14 |
| 12-24 | Single chopper used as both modulator and demodulator | 12-15 |
| 12-25 | Digital-to-voltage converter | 12-15 |
| 12-26 | Digital-to-voltage converter | 12-16 |
| 12-27 | Digital-to-voltage converter | 12-17 |
| 12-28 | Digital-to-voltage converter | 12-17 |
| 12-29 | Servomechanism for digital-to-analog conversion | 12-18 |

LIST OF ILLUSTRATIONS

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 13-1 | Symbolic representation of a diode | 13-1 |
| 13-2 | Volt-ampere curve of a diode cathode temperature constant | 13-2 |
| 13-3 | Diode used as a rectifier | 13-2 |
| 13-4 | Symbolic representation of a triode | 13-3 |
| 13-5 | Plate characteristics of a triode | 13-3 |
| 13-6 | Triode amplifier | 13-3 |
| 13-7 | Graphical analysis of a triode amplifier | 13-4 |
| 13-8 | Linear equivalent circuit of a triode | 13-4 |
| 13-9 | Alternate equivalent circuit of a triode | 13-5 |
| 13-10 | Symbolic representation of a beam-power tube and a pentode | 13-6 |
| 13-11 | Plate characteristics of a beam-power tube with constant screen voltage | 13-6 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 13-12 | Plate characteristics of a pentode with constant suppressor and screen voltages | 13-6 |
| 13-13 | Equivalent circuit of a simple plate-loaded amplifier | 13-7 |
| 13-14 | Series tube amplifier | 13-8 |
| 13-15 | Cascode amplifier circuit | 13-9 |
| 13-16 | Cathode follower | 13-9 |
| 13-17 | Equivalent circuit of a cathode follower and its manipulation | 13-10 |
| 13-18 | White cathode follower | 13-11 |
| 13-19 | Thevenin equivalent circuit of White cathode follower | 13-11 |
| 13-20 | Two-tube cathode follower | 13-11 |
| 13-21 | Plate-and-cathode-loaded amplifier | 13-13 |
| 13-22 | Differential amplifier | 13-13 |
| 13-23 | Push-pull amplifier | 13-14 |
| 13-24 | Graphical analysis of a push-pull amplifier | 13-16 |
| 13-25 | Power amplifier with a-c supply | 13-17 |
| 13-26 | Waveforms of power amplifier in Fig. 13-25 | 13-17 |
| 13-27 | Voltage-divider-coupled d-c amplifier | 13-17 |
| 13-28 | Battery-coupled, d-c amplifier | 13-18 |
| 13-29 | D-c amplifier with single supply voltage | 13-18 |
| 13-30 | D-c amplifier with gas-discharge tube coupling | 13-19 |
| 13-31 | Drift-compensated d-c amplifier | 13-20 |
| 13-32 | Drift-compensated d-c amplifier | 13-20 |
| 13-33 | Drift-compensated d-c amplifier | 13-20 |
| 13-34 | Use of a-c amplifier to replace d-c amplifier | 13-21 |
| 13-35 | Two-stage a-c amplifier (resistance-capacitance coupled) | 13-21 |
| 13-36 | Equivalent circuit of two-stage a-c amplifier in Fig. 13-35 .. | 13-22 |
| 13-37 | Simplified equivalent circuits of first stage of Fig. 13-35 | 13-22 |
| 13-38 | Frequency response of single-stage a-c coupled amplifier using resistance-capacitance coupling | 13-24 |
| 13-39 | Single-stage a-c coupled amplifier with voltage feedback | 13-25 |
| 13-40 | A-c servo amplifier | 13-29 |
| 13-41 | Typical thyatron control characteristics | 13-30 |
| 13-42 | Half-wave thyatron amplifier | 13-31 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 13-43 | Waveforms of circuit in Fig. 13-42 | 13-32 |
| 13-44 | Control of a thyatron by means of a phase-variable a-c signal | 13-33 |
| 13-45 | Plate and load connections of typical thyatron amplifiers | 13-33 |
| 13-46 | Load voltage and current supplied by a single-phase full-wave thyatron amplifier to a highly inductive load | 13-34 |
| 13-47 | Full-wave rectifier with typical L-C filter | 13-35 |
| 13-48 | Bridge rectifier with typical L-C filter | 13-35 |
| 13-49 | Block diagram of regulated power supply | 13-36 |
| 13-50 | Circuit schematic of series regulator | 13-37 |
| 13-51 | Temperature-limited diode amplifier | 13-38 |
| 13-52 | Typical characteristics of Type 1N137B silicon junction diode | 13-41 |
| 13-53 | Types of junction transistors | 13-43 |
| 13-54 | Typical collector characteristics | 13-44 |
| 13-55 | Approximate junction transistor model | 13-45 |
| 13-56 | Incremental transistor models | 13-45 |
| 13-57 | Models of h-parameter transistor | 13-47 |
| 13-58 | I_{co} temperature dependence | 13-48 |
| 13-59 | Variation of h-parameter with bias and temperature | 13-48 |
| 13-60 | Basic transistor amplifier configurations | 13-51 |
| 13-61 | Typical amplifier circuits | 13-52 |
| 13-62 | Push-pull Class B power-amplifier circuits | 13-53 |
| 13-63 | Typical biasing circuits | 13-54 |
| 13-64 | Two-stage transistor feedback amplifier | 13-55 |
| 13-65 | Collector characteristics of typical power transistor | 13-57 |
| 13-66 | Simple single-core magnetic amplifier | 13-59 |
| 13-67 | Simplified B-H characteristic of core material | 13-59 |
| 13-68 | Waveforms of source voltage, flux density, and load current over one cycle | 13-60 |
| 13-69 | Curves showing relationship between average load current I_1 , firing angle α , and control voltage E_c | 13-61 |
| 13-70 | Ramey circuit | 13-62 |
| 13-71 | Doubler circuit | 13-63 |
| 13-72 | Half-wave magnetic servo amplifier bridge circuit | 13-64 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 13-73 | Waveforms of load current during a transient change in output from minimum to maximum output (resistive load) | 13-65 |
| 13-74 | (Left) Power output vs total weight of 60-cycle standard self-saturating magnetic amplifiers. (Right) Comparison of power output vs reactor weight at 60 cps and 400 cps | 13-68 |
| 13-75 | Control characteristics for $N^2/R = 1.0$ zero bias current | 13-69 |
| 13-76 | Power gain per cycle for self-saturating magnetic amplifiers with several core materials in single and three-phase bridge circuits; with d-c output and 60-cycle supply | 13-69 |
| 13-77 | A rotary electric amplifier | 13-74 |
| 13-78 | Types of excitation in d-c rotary amplifiers | 13-75 |
| 13-79 | Block diagram representation of multifield d-c rotary amplifier | 13-78 |
| 13-80 | Types of control circuits used in rotary electric amplifiers | 13-86 |
| 13-81 | Typical saturation curves for a rotary electric amplifier | 13-86 |
| 13-82 | Speed control using single-sided single-stage relay amplifier | 13-88 |
| 13-83 | Speed control using single-sided cascade relay amplifier | 13-89 |
| 13-84 | Position control illustrating use of double-sided single-stage phase-sensitive relay amplifier | 13-90 |
| 13-85 | Position control with single-stage relay amplifier showing rate compensation or anticipation | 13-91 |
| 13-86 | Position control illustrating use of single-stage phase-sensitive relay amplifier | 13-93 |
| 13-87 | Position control illustrating use of two-stage phase-sensitive relay amplifier | 13-98 |
| 13-88 | Static characteristics of idealized relay | 13-99 |
| 13-89 | Error response of contactor servo with large input change | 13-100 |
| 13-90 | Circuit for measuring relay pull-in and drop-out time | 13-101 |
| 13-91 | Typical waveforms observed during relay test | 13-102 |
| 13-92 | Arc suppression circuit | 13-103 |
| 13-93 | Nomograph and equations for use in calculating the component values for an arc suppression circuit | 13-104 |
| 13-94 | Three-way spool-valve amplifier | 13-107 |
| 13-95 | Single nozzle-baffle amplifier | 13-108 |
| 13-96 | Four-way spool-valve amplifier | 13-108 |
| 13-97 | Four-way spool-valve amplifier with open center | 13-109 |
| 13-98 | Nozzle-baffle amplifier with balanced load | 13-109 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 13-99 | Sliding-plate valve amplifier | 13-109 |
| 13-100 | Amplifier with position feedback by means of moving valve sleeve | 13-110 |
| 13-101 | Amplifier with position feedback by means of linkage | 13-111 |
| 13-102 | Amplifier with feedback by means of force-balance system | 13-111 |
| 13-103 | Oil-gear rotary hydraulic amplifier with radial pistons | 13-112 |
| 13-104 | Constant-speed rotary hydraulic amplifier with axial pistons | 13-113 |
| 13-105 | Ball-and-piston type rotary hydraulic amplifier | 13-114 |
| 13-106 | Four-way spool valve, all pressures measured above sump pressure | 13-115 |
| 13-107 | Four-way spool valve, nondimensional plot of load pressure vs. spool displacement for zero load flow | 13-116 |
| 13-108 | Four-way spool valve, nondimensional plot of load pressure vs. spool displacement for zero load flow | 13-117 |
| 13-109 | Four-way spool valve, nondimensional plot of load flow vs. spool displacement for zero underlap and zero radial clearance ($b = X_o = 0$) | 13-117 |
| 13-110 | Four-way spool valve, nondimensional plot of load flow vs. spool displacement for zero underlap ($X_o = 0$) and finite radial clearance ($b \neq 0$) | 13-118 |
| 13-111 | Four-way spool valve, nondimensional plot of load flow vs. spool displacement for $\gamma = 0.5$ | 13-118 |
| 13-112 | Four-way spool valve, equivalent hydraulic-source representation | 13-118 |
| 13-113 | Four-way spool valve nondimensional plot of flow gain vs. load pressure for zero radial clearance ($b = 0$) | 13-119 |
| 13-114 | Four-way spool valve with negligible radial clearance, nondimensional plot of internal conductance vs. load pressure .. | 13-120 |
| 13-115 | Four-way spool valve, nondimensional plot of load flow vs. differential load pressure for balanced load | 13-121 |
| 13-116 | Equivalent circuit of four-way spool-valve amplifier with balanced load | 13-121 |
| 13-117 | Simplified equivalent circuits of four-way spool-valve amplifier with balanced load | 13-121 |
| 13-118 | Geometry of spool valve with radial clearance and underlap .. | 13-122 |
| 13-119 | Four-way spool valve, nondimensional plot of flow gain vs. load pressure for $\beta = 0.5$ | 13-122 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 13-120 | Four-way spool valve, nondimensional plot of flow gain vs. load pressure for $\beta = 1.0$ | 13-124 |
| 13-121 | Four-way spool valve, nondimensional plot of flow gain vs. load pressure for $\beta = 5$ | 13-124 |
| 13-122 | Plot of $\frac{\left \frac{\delta Q}{\delta P} \right }{\left[\frac{C\pi d X_o}{2\sqrt{P_s}} \right]}$ vs $\frac{P}{P_s}$ with $\frac{x}{X_o}$ as parameter with $\beta = 0.5, 1.0$, and 5.0 | 13-125 |
| 13-123 | Four-way spool valve with finite radial clearance, plot of load flow vs spool displacement for $\frac{P_s}{P_s} = \frac{100 \text{ (lb/in.}^2\text{)}}{350 \text{ (lb/in.}^2\text{)}} = \text{constant}$ | 13-128 |
| 13-124 | Geometry of orifice plate valve with underlap and clearance | 13-129 |
| 13-125 | Double nozzle-baffle valve with balanced piston load | 13-130 |
| 13-126 | Double nozzle-baffle valve, nondimensional plot of pressure gain vs load pressure | 13-132 |
| 13-127 | Double nozzle-baffle valve, nondimensional plot of internal resistance vs load pressure | 13-133 |
| 13-128 | Double nozzle-baffle valve, nondimensional plot of load pressure vs baffle displacement for zero load flow | 13-134 |
| 13-129 | Double nozzle-baffle valve with balanced piston load, nondimensional plot of load flow vs differential load pressure .. | 13-135 |
| 13-130 | Pressure-control valve amplifier (single-sided amplifier with pressure compensation) | 13-136 |
| 13-131 | Pressure-control valve amplifier with balanced load, dimensional plot of differential load pressure vs load flow | 13-136 |
| 13-132 | Rotary pump, typical dimensional plot of load flow vs load pressure | 13-137 |
| 13-133 | Rotary hydraulic pump | 13-138 |
| 13-134 | Four-way spool-valve amplifier and load; equivalent hydraulic-circuit representation | 13-139 |
| 13-135 | Equivalent circuit representation of load having mass and opposing force | 13-146 |
| 13-136 | Equivalent circuit representation of load having mass and spring and opposing force | 13-147 |
| 13-137 | Equivalent circuit representation of load having mass, spring, viscous damping, and opposing force | 13-148 |
| 13-138 | Piston with unequal working areas | 13-148 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|--|-------------|
| 13-139 | Equivalent circuit representation of load having mass, friction, compliance, and opposing force | 13-148 |
| 13-140 | Four-way spool-valve amplifier and load block diagram — hydraulic circuit shown in Fig. 13-134 — load has spring, mass, and dashpot | 13-149 |
| 13-141 | Four-way spool-valve amplifier; simplified block diagram derived from Fig. 13-140 | 13-149 |
| 13-142 | Equivalent hydraulic-circuit representation of rotary amplifier (pump) with load having hydraulic compressibility, leakage, and attached mass, spring, and viscous friction | 13-150 |
| 13-143 | Commercially available electrohydraulic servo valves | 13-152 |
| 13-144 | Block diagram of electric amplifier, torque motor, and first- and second-stage hydraulic amplifier | 13-154 |
| 13-145 | Relative compressibility coefficient as a function of pressure and amount of entrained gas | 13-157 |
| 13-146 | Simplified circuit diagram for illustrative example | 13-161 |
| 13-147 | Simplified block diagram for illustrative example | 13-162 |
| 13-148 | Schematic of sliding-plate valve | 13-164 |
| 13-149 | Schematic of conical-plug type three-way valve | 13-165 |
| 13-150 | Schematic of nozzle-baffle three-way valve | 13-166 |
| 13-151 | Plot of restriction factor versus pressure ratio for flow of compressible fluid through an orifice | 13-166 |
| 13-152 | Nondimensional plot of theoretical load flow versus load pressure for open-center (under-lapped) three-way valve .. | 13-168 |
| 13-153 | Nondimensional plot of internal shunt conductance versus valve displacement for conical-plug valve | 13-170 |
| 13-154 | Nondimensional plot of flow gain versus valve displacement for conical-plug valve | 13-170 |
| 13-155 | Nondimensional plot of internal shunt conductance versus baffle opening for nozzle-baffle valve | 13-171 |
| 13-156 | Nondimensional plot of gain versus baffle opening for nozzle-baffle valve | 13-171 |
| 13-157 | Nondimensional plot of load pressure versus baffle opening for nozzle-baffle valve | 13-172 |
| 13-158 | Plots of (1) discharge coefficient versus baffle opening; and (2) discharge coefficient-baffle opening product versus baffle opening for typical exhaust nozzle in nozzle-baffle valve | 13-172 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 13-159 | Plot of pressure ratio versus exhaust-to-supply orifice area ratio for nozzle-baffle valve | 13-173 |
| 13-160 | Equivalent-circuit representation for pneumatic amplifier (three-way valve) with spring-opposed ram load | 13-174 |
| 13-161 | Frequency-response curve for a four-way pneumatic servo-mechanism with 1000 psi supply pressure (plate-valve and piston combination with position feedback by electrical means) | 13-176 |
| 13-162 | Ball-disc integrator | 13-181 |
| 13-163 | Cone-and-disc amplifier | 13-182 |
| 13-164 | Two forms of double capstan amplifier | 13-183 |
| 13-165 | Ideal and actual characteristics of mechanical amplifiers | 13-184 |
| 13-166 | Typical speed-torque characteristics of variable-speed-output mechanical amplifier | 13-184 |
| 13-167 | Schematic of working parts of integrator type mechanical amplifier | 13-185 |
| 13-168 | Block diagram of capstan amplifier, with load inertia J and load torque T_L | 13-187 |
| 13-169 | Capstan amplifier with electrical input | 13-190 |
| 13-170 | Block diagram of integrator-type amplifier with inertia load and external feedback | 13-192 |
| 13-171 | Solenoid-operated clutch | 13-193 |
| 13-172 | Solenoid-operated clutches | 13-194 |

LIST OF ILLUSTRATIONS

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 14-1 | Rotary electric amplifier circuits for control of d-c motor | 14-2 |
| 14-2 | High-vacuum tube circuit for control of d-c motor | 14-4 |
| 14-3 | Gas-filled tube circuits for control of d-c motors | 14-5 |
| 14-4 | Magnetic amplifier circuit for control of series motor | 14-6 |
| 14-5 | Magnetic amplifier circuit for unidirectional control of shunt motor | 14-7 |
| 14-6 | Relay amplifier circuits for control of d-c motors | 14-8 |
| 14-7 | Static torque-speed characteristics of armature-controlled d-c motor | 14-9 |
| 14-8 | Static torque-speed characteristics of field-controlled d-c motor, constant armature current | 14-10 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 14-9 | Static torque-speed characteristics of field-controlled d-c motor, constant armature voltage | 14-11 |
| 14-10 | Static torque-speed characteristics of series motor | 14-11 |
| 14-11 | Block diagram for armature-controlled d-c motor | 14-12 |
| 14-12 | Block diagram for field-controlled d-c motor, constant armature current | 14-13 |
| 14-13 | Block diagram for field-controlled d-c motor, constant armature terminal voltage, incremental behavior | 14-15 |
| 14-14 | D-c torque motor, permanent-magnet type | 14-22 |
| 14-15 | Characteristics of d-c torque motor | 14-23 |
| 14-16 | Amplifier control circuit for shaded-pole motor | 14-25 |
| 14-17 | Amplifier control circuit for 2-phase servomotor | 14-26 |
| 14-18 | Schematic of a transistor amplifier used with a 2-phase servomotor | 14-27 |
| 14-19 | Full-wave output-stage circuit of a magnetic servo amplifier having inherent dynamic-braking properties: (A) with two separate control winding elements N'_c, N''_c ; (B) with series-connected d-c control winding elements N'_c, N''_c ; (C) with series-connected a-c control winding elements N'_c, N''_c | 14-28 |
| 14-20 | A-c motor control (1 phase fixed) | 14-29 |
| 14-21 | A-c motor control (both phases variable) | 14-30 |
| 14-22 | Relay servo using wound-shading-coil or 2-phase motors | 14-31 |
| 14-23 | Relay servo using capacitor-start-and-run motor | 14-31 |
| 14-24 | Torque-speed curves of a 2-phase servomotor | 14-32 |
| 14-25 | Damping factor for motor having torque-speed curves shown in Fig. 14-24 | 14-33 |
| 14-26 | Damping factors for some typical motors, taken with zero speed, $\Omega/\Omega_{\text{sync}} = 0$ | 14-33 |
| 14-27 | Torque gain of typical 2-phase servomotor versus per unit control voltage at different speeds | 14-33 |
| 14-28 | Illustrating the forward and backward sets of voltage derived from an unbalanced excitation for a 2-phase motor | 14-34 |
| 14-29 | Block diagrams of 2-phase motor and load | 14-35 |
| 14-30 | Torque-speed curves, constant-displacement hydraulic motor | 14-38 |
| 14-31 | Typical pneumatic capsule | 14-44 |
| 14-32 | Typical pneumatic bellows | 14-45 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 14-33 | Typical pneumatic diaphragm motor | 14-45 |
| 14-34 | Typical pneumatic ram | 14-45 |
| 14-35 | Typical pneumatic piston | 14-45 |
| 14-36 | Typical piston and cylinder with stabilizing capillaries and tanks | 14-46 |
| 14-37 | Simplified packing gland | 14-46 |
| 14-38 | Cross section of magnetic-particle clutch | 14-48 |
| 14-39 | Cross section of magnetic-particle clutch | 14-49 |
| 14-40 | Cross section of magnetic-particle clutch | 14-50 |
| 14-41 | Pictorial diagram of push-pull arrangement of dual-clutch servomechanism | 14-50 |
| 14-42 | Typical static torque-speed curves for magnetic-particle clutch | 14-50 |
| 14-43 | Torque vs coil current of single magnetic-particle clutch | 14-51 |
| 14-44 | Block diagram for magnetic-particle clutch and preamplifier, inertia load | 14-51 |
| 15-1 | Most important parts of a gear | 15-2 |
| 15-2 | Types of gears | 15-4 |
| 15-3 | Methods of eliminating backlash | 15-14 |
| 15-4 | Geared differentials | 15-16 |
| 15-5 | Differential lever | 15-17 |
| 15-6 | Typical linkages | 15-18 |
| 15-7 | Metal tape and sheaves | 15-20 |
| 15-8 | Coupling types | 15-23 |
| 15-9 | Square-parallel stock key | 15-25 |
| 15-10 | Woodruff key — SAE standard | 15-25 |
| 15-11 | Taper pin | 15-28 |
| 15-12 | SAE square-spline hub | 15-28 |
| 15-13 | Roller bearings | 15-30 |
| 15-14 | Flange-type ball bearing | 15-31 |
| 15-15 | Torque-tube bearing | 15-31 |
| 15-16 | Double-row ball bearing | 15-31 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 16-1 | Viscous-coupled inertia damper | 16-16 |
| 16-2 | Comparison of loop transfer functions | 16-17 |
| 16-3 | Comparison of torque constants | 16-17 |
| 16-4 | Functional block diagram of servo data repeater | 16-18 |
| 16-5 | Pictorial diagram of damper-motor | 16-19 |
| 16-6 | Operational block diagram for servo data repeater | 16-21 |
| 16-7 | Servo data repeater | 16-22 |
| 16-8 | Transient response of servo data repeater to various steps .. | 16-22 |
| 16-9 | Cross section of clutch-damper | 16-23 |
| 16-10 | Dual-mode package | 16-24 |
| 16-11 | Functional block diagram of dual-mode servo data repeater | 16-24 |
| 16-12 | Comparison of large-step (53°) servo response for single-mode and dual-mode operation | 16-25 |
| 17-1 | Simplified functional block diagram of servo system for controlling M33 tracking-radar antenna in elevation | 17-2 |
| 17-2 | Operational block diagram of servo system for controlling M33 tracking-radar antenna in elevation | 17-4 |
| 17-3 | Power control system for M-38 Fire-Control System | 17-6 |
| 17-4 | Operational block diagram of fine-speed portion of power control system for M-38 Fire-Control System | 17-8 |
| 17-5 | Simplifier block diagram of power control system for M-38 Fire-Control System | 17-9 |
| 17-6 | Final block diagram of power control system for M-38 Fire-Control System | 17-9 |
| 18-1 | Gear pump | 18-2 |
| 18-2 | Vane pump | 18-3 |
| 18-3 | Typical auxiliaries used with hydraulic transmission | 18-5 |
| 18-4 | Auxiliaries used in control-valve and hydraulic-ram system | 18-5 |
| 18-5 | Ball check valve or relief valve | 18-6 |
| 18-6 | Double-acting relief valve | 18-7 |
| 18-7 | Pressure-regulating valve | 18-7 |
| 18-8 | Throttling pressure regulator | 18-8 |
| 18-9 | Gravity accumulator | 18-8 |

LIST OF ILLUSTRATIONS (cont)

| <i>Fig. No.</i> | <i>Title</i> | <i>Page</i> |
|-----------------|---|-------------|
| 18-10 | Hydropneumatic accumulators | 18-9 |
| 18-11 | Unloading valve | 18-11 |
| 18-12 | Packing gland | 18-12 |
| 18-13 | Applications of O-rings | 18-13 |
| 18-14 | Shaft seals | 18-13 |
| 18-15 | Face seal | 18-14 |
| 18-16 | High-pressure seal | 18-14 |
| 18-17 | Switching limit-stop system | 18-16 |
| 18-18 | Electrical limit-stop system | 18-17 |
| 18-19 | Mechanical limit-stop system | 18-18 |
| 18-20 | Positive stop | 18-19 |
| 18-21 | Stop with lugged washers | 18-19 |
| 18-22 | Spring and buffer stop | 18-19 |
| 18-23 | Buffer stop in hydraulic cylinder | 18-19 |
| 19-1 | Vibration isolation mounting | 19-8 |
| 19-2 | Frequency response of vibration isolation mounting | 19-9 |
| 19-3 | Analytical model of shock-mounted device | 19-12 |
| 20-1 | Moments of inertia about the principal axes of cylinders one inch long | 20-3 |
| 20-2 | Relation between Saybolt seconds universal and stokes | 20-6 |
| 20-3 | Viscosity versus temperature for some oils and fluids | 20-7 |
| 20-4 | Cantilever spring | 20-11 |

LIST OF TABLES

| <i>Table No.</i> | <i>Title</i> | <i>Page</i> |
|-------------------------|--|--------------------|
| 3-1 | Laplace transform pairs | 3-14 |
| 3-2 | Block diagram symbols | 3-18 |
| 3-3 | Parameters of trapezoids | 3-30 |
| 3-4 | Error coefficients in terms of open-loop function $C(s)/E(s)$ | 3-35 |
| 5-1 | Properties of M and ϕ contours | 5-12 |
| 6-1 | Results of lead compensation example | 6-6 |
| 6-2 | Results of compensation using 45° phase margin | 6-10 |
| 7-1 | Common performance indices | 7-46 |
| 8-1 | Table of definite integrals | 8-4 |
| 8-2 | The minimum ITAE standard forms, zero-displacement-error systems | 8-5 |
| 8-3 | The minimum ITAE standard forms, zero-velocity-error systems | 8-5 |
| 9-1 | Laplace and z transform pairs | 9-4 |
| 9-2 | Modified z transforms | 9-13 |

LIST OF TABLES

| <i>Table No.</i> | <i>Title</i> | <i>Page</i> |
|------------------|--|-------------|
| 11-1 | Load compensation by the addition of both series and parallel resistors | 11-8 |
| 11-2 | Load compensation by the addition of a parallel resistor | 11-8 |
| 11-3 | Approximation of nonlinear functions through slider loading and series resistors | 11-12 |
| 11-4 | Functional classification of synchros | 11-19 |
| 11-5 | Limits of tolerances for military synchros (115 volts, 60 cps) | 11-30 |
| 11-6 | Limiting values for 115-volt, 400-cps synchros | 11-33 |
| 11-7 | Limiting values for 26-volt, 400-cps synchros | 11-36 |
| 11-8 | Typical characteristics of commercial resolvers | 11-40 |
| 11-9 | Typical performance characteristics of microsyn units | 11-47 |
| 11-10 | Sample data for drag-cup a-c tachometer generators | 11-59 |
| 11-11 | Binary numbers | 11-80 |
| 11-12 | Cyclic code and its decimal and binary equivalents | 11-83 |

LIST OF TABLES

| <i>Table No.</i> | <i>Title</i> | <i>Page</i> |
|------------------|--|-------------|
| 13-1 | Symbols | 13-1 |
| 13-2 | Relations between incremental parameters | 13-46 |
| 13-3 | Typical parameter variation for a low-power transistor | 13-49 |
| 13-4 | Comparison of typical design information on silicon and germanium power transistors | 13-56 |
| 13-5 | Typical characteristics of magnetic amplifiers | 13-67 |
| 13-6 | Dynamic characteristics of some basic configurations of rotary electric amplifiers | 13-81 |
| 13-7 | Inductances and sensitivities associated with the windings of rotary electric amplifiers | 13-84 |
| 13-8 | Some typical values of parameters for rotary electric amplifiers | 13-87 |
| 13-9 | Comparison of commercial relays | 13-94 |
| 13-10 | Methods of increasing relay sensitivity | 13-99 |
| 13-11 | Methods of increasing contact rating | 13-99 |
| 13-12 | Methods of increasing relay response speed | 13-101 |
| 13-13 | Methods of increasing R/δ | 13-105 |
| 13-14 | Classification of hydraulic amplifiers | 13-106 |
| 13-15 | Four-way spool valve with appreciable radial clearance — equivalent-source flow gain and conductance parameters | 13-123 |
| 13-16 | Transfer functions of four-way spool-valve amplifier and load | 13-140 |
| 13-17 | Transfer functions of rotary amplifier and load | 13-142 |
| 13-18 | Circuit parameters for rotary hydraulic amplifier and load | 13-144 |
| 13-19 | Dynamic and static characteristics of commercially available electrohydraulic servo control valves | 13-151 |
| 13-20 | Friction factors | 13-156 |
| 13-21 | Elasticity factors for pipe | 13-157 |
| 13-22 | Orifice area equations for pneumatic three-way valves | 13-167 |
| 13-23 | Minimum tubing lengths for linear flow | 13-175 |
| 13-24 | Calculated time constants for a three-way conical-plug valve | 13-179 |
| 13-25 | Characteristics of VSD units | 13-188 |
| 13-26 | Some typical values of mechanical amplifier parameters | 13-191 |
| 13-27 | Characteristics of solenoid-operated clutch | 13-195 |

LIST OF TABLES

| <i>Table No.</i> | <i>Title</i> | <i>Page</i> |
|------------------|--|-------------|
| 14-1 | Typical values of d-c motor parameters | 14-21 |
| 14-2 | Conversion factors and units | 14-24 |
| 14-3 | Comparison of typical motors | 14-36 |
| 14-4 | Analogous parameters of constant-displacement hydraulic motor and electric shunt motor | 14-39 |
| 14-5 | References to tables listing hydraulic-transmission transfer functions | 14-40 |
| 14-6 | Transmission data — hydraulic pumps | 14-40 |
| 14-7 | Transmission data — hydraulic motors | 14-41 |
| 14-8 | Transmission data — fluid | 14-41 |
| 14-9 | Transmission data — general | 14-42 |
| 14-10 | Comparison of predicted and measured parameters | 14-42 |
| 14-11 | Properties of a typical line of magnetic-particle clutches | 14-52 |
| 14-12 | Typical time constants | 14-53 |
| 14-13 | Summary of clutches using various magnetic-particle mixtures | 14-54 |
| 15-1 | Values of deformation factor C | 15-6 |
| 15-2 | Maximum error in action between gears as a function of class | 15-7 |
| 15-3 | Maximum error in action between gears as a function of pitch line velocity | 15-7 |
| 15-4 | Values of tooth form factor y for various tooth forms | 15-9 |
| 15-5 | Values of safe static bending stress s_t | 15-10 |
| 15-6 | Values of load-stress factor K for various materials | 15-11 |
| 15-7 | Values of load-stress factor K for hardened steel | 15-12 |
| 15-8 | Sheave diameters | 15-21 |
| 15-9 | Dimensions for square-parallel stock keys | 15-26 |
| 15-10 | Dimensions of some Woodruff keys, keyslots, and keyways | 15-27 |
| 15-11 | Dimensions of taper pins | 15-28 |
| 15-12 | Dimensions of commercial unshielded inch-type ball bearings with retainers | 15-32 |

LIST OF TABLES (cont)

| <i>Table No.</i> | <i>Title</i> | <i>Page</i> |
|------------------|--|-------------|
| 15-13 | Dimensions of commercial unshielded flange-type bearings | 15-33 |
| 15-14 | Dimensions of torque-tube bearings | 15-33 |
| 16-1 | Comparison of servomotors | 16-7 |
| 17-1 | Nomenclature for M-38 power control system | 17-10 |
| 19-1 | Heat liberated by control devices | 19-3 |
| 19-2 | Emissivity of various surfaces | 19-4 |
| 19-3 | Values of [H_F in Btu/(hr) (ft ²) (°F)] | 19-5 |
| 19-4 | Thermal conductivity of metals and alloys at 212°F | 19-6 |
| 19-5 | Conversion factors for coefficients of heat transfer | 19-7 |
| 19-6 | Typical vibration and shock values encountered in ordnance equipment | 19-11 |
| 20-1 | Relations between units of mass moment of inertia | 20-2 |
| 20-2 | Relations between torque, moment of inertia, and acceleration | 20-2 |
| 20-3 | Relations between units of absolute viscosity | 20-5 |
| 20-4 | Coefficients of friction for various metals | 20-8 |
| 20-5 | Torsional elastic properties of some spring materials | 20-10 |
| 20-6 | Permissible load and spring scale of helical compression springs | 20-12 |
| 20-7 | Tensile elastic properties of some spring materials | 20-14 |
| 20-8 | Elastic properties of tubing material | 20-16 |

TABLE OF CONTENTS

| <i>Paragraph</i> | | <i>Page</i> |
|--|---|-------------|
| CHAPTER 1 | | |
| PROPERTIES OF FEEDBACK CONTROL SYSTEMS | | |
| 1-1 | OBJECTIVES OF A FEEDBACK CONTROL SYSTEM | 1-1 |
| 1-2 | OPEN-LOOP VS CLOSED-LOOP SYSTEM CHARACTERISTICS | 1-2 |
| 1-3 | STABILITY AND DYNAMIC RESPONSE | 1-2 |
| 1-4 | TERMINOLOGY OF FEEDBACK CONTROL SYSTEMS | 1-3 |
| CHAPTER 2 | | |
| DYNAMIC RESPONSE | | |
| 2-1 | INTRODUCTION | 2-1 |
| 2-2 | LINEARIZATION | 2-1 |
| 2-3 | TRANSIENT RESPONSE | 2-5 |
| 2-4 | FREQUENCY RESPONSE | 2-7 |
| 2-5 | FORCED RESPONSE | 2-8 |
| 2-6 | STOCHASTIC INPUTS | 2-8 |
| CHAPTER 3 | | |
| METHODS OF DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS | | |
| 3-1 | THE DIFFERENTIAL EQUATIONS | 3-1 |
| 3-2 | FACTORING AND CHARACTERISTIC PARAMETERS OF RESPONSE MODES | 3-2 |
| 3-2.1 | FACTORING | 3-2 |
| 3-2.2 | CHARACTERISTIC PARAMETERS OF RESPONSE MODES | 3-4 |
| 3-2.3 | First Order | 3-4 |
| 3-2.4 | Second Order | 3-5 |
| 3-2.5 | Third Order | 3-5 |
| 3-2.6 | Fourth Order | 3-5 |
| 3-3 | THE CONVOLUTION INTEGRAL | 3-11 |
| 3-4 | LAPLACE AND FOURIER TRANSFORMS | 3-12 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 3 (cont) | |
| 3-4.1 | GENERAL | 3-12 |
| 3-4.2 | THEOREMS | 3-12 |
| 3-4.3 | SOLUTION OF DIFFERENTIAL EQUATIONS | 3-13 |
| 3-4.4 | FREQUENCY RESPONSE | 3-17 |
| 3-5 | BLOCK DIAGRAMS AND SIGNAL-FLOW GRAPHS .. | 3-17 |
| 3-5.1 | BLOCK DIAGRAMS | 3-17 |
| 3-5.2 | SIGNAL-FLOW GRAPHS | 3-25 |
| 3-6 | APPROXIMATE NUMERICAL AND GRAPHICAL METHODS OF DETERMINING TRANSIENT RE- SPONSE | 3-34 |
| 3-7 | ERROR COEFFICIENTS FOR DETERMINING RE- SPONSE TO AN ARBITRARY INPUT | 3-34 |
| 3-8 | RESPONSE TO STATIONARY STOCHASTIC INPUTS | 3-36 |
| 3-9 | USE OF ANALOG COMPUTERS FOR SIMULATION | 3-39 |
| | CHAPTER 4 | |
| | STABILITY OF FEEDBACK CONTROL SYSTEMS | |
| 4-1 | INTRODUCTION | 4-1 |
| 4-2 | ROUTH CRITERION | 4-2 |
| 4-3 | NYQUIST CRITERION | 4-4 |
| 4-4 | ROOT-LOCUS METHOD | 4-7 |
| | CHAPTER 5 | |
| | GAIN DETERMINATION | |
| 5-1 | PERFORMANCE CRITERIA AND DEFINITIONS | 5-1 |
| 5-1.1 | GENERAL | 5-1 |
| 5-1.2 | GAIN | 5-1 |
| 5-1.3 | VELOCITY CONSTANT | 5-1 |
| 5-1.4 | ACCELERATION CONSTANT | 5-1 |
| 5-1.5 | TORQUE CONSTANT | 5-2 |
| 5-1.6 | STATIC ACCURACY | 5-2 |
| 5-1.7 | BANDWIDTH | 5-2 |
| 5-1.8 | PEAK MAGNITUDE | 5-3 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 5 (cont) | <i>Page</i> |
|------------------|--|-------------|
| 5-2 | POLAR-PLANE REPRESENTATION | 5-3 |
| 5-2.1 | GENERAL | 5-3 |
| 5-2.2 | DIRECT POLAR PLANE | 5-3 |
| 5-2.3 | INVERSE POLAR PLANE | 5-4 |
| 5-3 | EXACT AND ASYMPTOTIC-LOGARITHMIC REPRESENTATIONS | 5-4 |
| 5-3.1 | GENERAL | 5-4 |
| 5-3.2 | SEPARATE MAGNITUDE AND PHASE PLOTS | 5-5 |
| 5-3.3 | Magnitude Curves | 5-5 |
| 5-3.4 | Phase-Angle Curves | 5-7 |
| 5-3.5 | GAIN-PHASE PLANE | 5-10 |
| 5-4 | CLOSED-LOOP RESPONSE DETERMINATION | 5-10 |
| 5-4.1 | GENERAL | 5-10 |
| 5-4.2 | POLAR-PLANE TECHNIQUE | 5-11 |
| 5-4.3 | GAIN-PHASE PLANE TECHNIQUE (NICHOLS CHART) | 5-13 |
| 5-4.4 | NONUNITY-FEEDBACK SYSTEMS | 5-15 |
| 5-5 | SETTING THE GAIN FOR A SPECIFIED M_p | 5-15 |
| 5-5.1 | GENERAL | 5-15 |
| 5-5.2 | POLAR-PLANE CONSTRUCTION | 5-16 |
| 5-5.3 | GAIN-PHASE PLANE CONSTRUCTION | 5-18 |
| 5-6 | APPROXIMATE PROCEDURES | 5-20 |
| 5-6.1 | PHASE MARGIN AND GAIN MARGIN | 5-20 |
| 5-6.2 | Phase Margin | 5-20 |
| 5-6.3 | Gain Margin | 5-20 |
| 5-6.4 | GENERAL COMMENTS ON THE PHASE-MARGIN CRITERION | 5-21 |
| 5-6.5 | APPROXIMATE CLOSED-LOOP RESPONSE | 5-21 |
| 5-7 | ROOT-LOCUS METHOD | 5-23 |
| 5-7.1 | GENERAL | 5-23 |
| 5-7.2 | PROPERTIES OF ROOTS IN THE s PLANE | 5-23 |
| 5-7.3 | First-Order Root | 5-23 |
| 5-7.4 | Second-Order Root | 5-23 |
| 5-7.5 | GAIN DETERMINATION IN THE s PLANE | 5-24 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 6 | <i>Page</i> |
|------------------|---|-------------|
| | COMPENSATION TECHNIQUES | |
| 6-1 | INTRODUCTION | 6-1 |
| 6-2 | RESHAPING LOCUS ON GAIN-PHASE PLANE | 6-2 |
| 6-2.1 | GENERAL | 6-2 |
| 6-2.2 | LAG COMPENSATION | 6-2 |
| 6-2.3 | LEAD COMPENSATION | 6-4 |
| 6-3 | PHASE-MARGIN AND ASYMPTOTIC METHODS | 6-6 |
| 6-3.1 | GENERAL | 6-6 |
| 6-3.2 | LAG COMPENSATION | 6-7 |
| 6-3.3 | LEAD COMPENSATION | 6-7 |
| 6-4 | FEEDBACK OR PARALLEL COMPENSATION | 6-10 |
| 6-5 | ALTERNATIVE DESIGN METHODS | 6-15 |
| 6-6 | TYPICAL COMPENSATION NETWORKS | 6-16 |
| 6-6.1 | D-C ELECTRIC | 6-16 |
| 6-6.2 | A-C ELECTRIC | 6-17 |
| 6-6.3 | MECHANICAL DAMPER | 6-19 |
| 6-6.4 | HYDRAULIC AMPLIFIER | 6-20 |
| 6-6.5 | PNEUMATIC CONTROLLER | 6-21 |
| | CHAPTER 7 PERFORMANCE EVALUATION | |
| 7-1 | RELATIONS BETWEEN FREQUENCY RESPONSE AND TRANSIENT RESPONSE | 7-1 |
| 7-1.1 | GENERAL | 7-1 |
| 7-1.2 | CLOSED-LOOP FREQUENCY RESPONSE FROM CLOSED-LOOP TRANSIENT RESPONSE | 7-1 |
| 7-1.3 | RELATIONS BETWEEN CLOSED-LOOP TRAN- SIENT RESPONSE AND CLOSED-LOOP POLE- ZERO CONFIGURATION | 7-17 |
| 7-1.4 | RELATIONS BETWEEN OPEN-LOOP FREQUEN- CY RESPONSE AND CLOSED-LOOP TRANSIENT RESPONSE | 7-21 |
| 7-2 | ERROR COEFFICIENTS | 7-44 |
| 7-3 | PERFORMANCE INDICES | 7-45 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|---|---|-------------|
| CHAPTER 8 | | |
| OPTIMIZATION METHODS FOR TRANSIENT AND STOCHASTIC INPUTS | | |
| 8-1 | CRITERIA OF PERFORMANCE | 8-1 |
| 8-2 | OPTIMUM SYNTHESIS OF FIXED-CONFIGURATION SYSTEMS | 8-2 |
| 8-2.1 | TRANSIENT INPUTS | 8-2 |
| 8-2.2 | STATIONARY STOCHASTIC INPUTS | 8-6 |
| 8-3 | OPTIMUM SYNTHESIS OF FREE-CONFIGURATION SYSTEMS WITH STATIONARY STOCHASTIC INPUTS | 8-8 |
| 8-4 | LIMITATIONS AND APPLICATION PROBLEMS | 8-11 |
| CHAPTER 9 | | |
| SAMPLE-DATA SYSTEMS | | |
| 9-1 | GENERAL THEORY | 9-1 |
| 9-2 | THE z TRANSFORM AND THE w TRANSFORM | 9-4 |
| 9-2.1 | THE z TRANSFORM | 9-4 |
| 9-2.2 | THE w TRANSFORM | 9-5 |
| 9-3 | OPERATIONAL METHODS | 9-6 |
| 9-3.1 | GENERAL | 9-6 |
| 9-3.2 | BASIC RELATIONS OF SAMPLED FUNCTIONS | 9-6 |
| 9-3.3 | ADDITIONAL PROPERTIES OF SAMPLED FUNCTIONS | 9-8 |
| 9-4 | DESIGN TECHNIQUES | 9-9 |
| 9-5 | PERFORMANCE EVALUATION | 9-10 |
| CHAPTER 10 | | |
| NONLINEAR SYSTEMS | | |
| 10-1 | INTRODUCTION | 10-1 |
| 10-2 | DESCRIBING FUNCTION PROCEDURES | 10-1 |
| 10-3 | PHASE-PLANE PROCEDURES | 10-8 |
| 10-4 | LIMITATIONS, COMPENSATION, AND OTHER METHODS | 10-13 |

TABLE OF CONTENTS

| <i>Paragraph</i> | CHAPTER 11 | <i>Page</i> |
|------------------|---|-------------|
| | SENSING ELEMENTS | |
| 11-1 | INTRODUCTION | 11-1 |
| 11-2 | POTENTIOMETERS | 11-1 |
| 11-2.1 | DESCRIPTION AND BASIC THEORY | 11-1 |
| 11-2.2 | Definition | 11-1 |
| 11-2.3 | Types of Potentiometers | 11-2 |
| 11-2.4 | Principle of Operation | 11-2 |
| 11-2.5 | Use | 11-3 |
| 11-2.6 | Construction Features | 11-3 |
| 11-2.7 | LINEAR POTENTIOMETERS | 11-4 |
| 11-2.8 | Types of Linearity | 11-4 |
| 11-2.9 | Effect of Load Impedance | 11-6 |
| 11-2.10 | NONLINEAR POTENTIOMETERS | 11-7 |
| 11-2.11 | APPLICATION FACTORS | 11-11 |
| 11-2.12 | Noise | 11-11 |
| 11-2.13 | Power Rating | 11-13 |
| 11-2.14 | Environmental Effects | 11-13 |
| 11-2.15 | Life | 11-14 |
| 11-2.16 | Mechanical Loading | 11-14 |
| 11-2.17 | Lubrication | 11-15 |
| 11-3 | ROTARY TRANSFORMERS | 11-15 |
| 11-3.1 | GENERAL DESCRIPTION | 11-15 |
| 11-3.2 | GENERAL CLASSIFICATIONS | 11-15 |
| 11-3.3 | Use in Positional Systems | 11-15 |
| 11-3.4 | Miscellaneous System Uses | 11-15 |
| 11-3.5 | General Functional Classification | 11-15 |
| 11-3.6 | General Unit Classification | 11-16 |
| 11-3.7 | SYNCHROS | 11-16 |
| 11-3.8 | Stator Construction | 11-16 |
| 11-3.9 | Rotor Construction | 11-16 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 11 (cont) | |
| 11-3.10 | Synchro Supply | 11-16 |
| 11-3.11 | Nomenclature | 11-16 |
| 11-3.12 | Other Methods of Nomenclature | 11-18 |
| 11-3.13 | Transmitter Characteristics | 11-18 |
| 11-3.14 | Receiver Characteristics | 11-19 |
| 11-3.15 | Transformer Characteristics | 11-20 |
| 11-3.16 | Differential Transmitter Characteristics | 11-22 |
| 11-3.17 | Differential Receiver Characteristics | 11-22 |
| 11-3.18 | Synchro Capacitors | 11-25 |
| 11-3.19 | Dual-Speed Synchro Systems | 11-25 |
| 11-3.20 | Zeroing | 11-25 |
| 11-3.21 | Zeroing a Transmitter (or Receiver) | 11-26 |
| 11-3.22 | Zeroing a Control Transformer | 11-26 |
| 11-3.23 | Zeroing a Differential Synchro Unit | 11-26 |
| 11-3.24 | Torque Relationships | 11-27 |
| 11-3.25 | Torque gradient | 11-27 |
| 11-3.26 | Performance prediction | 11-27 |
| 11-3.27 | Synchro Accuracy | 11-29 |
| 11-3.28 | Static errors | 11-29 |
| 11-3.29 | Dynamic errors | 11-29 |
| 11-3.30 | Military Specifications | 11-31 |
| 11-3.31 | INDUCTION RESOLVERS | 11-37 |
| 11-3.32 | Basic Operation | 11-37 |
| 11-3.33 | Design Principles | 11-37 |
| 11-3.34 | Applications | 11-38 |
| 11-3.35 | Booster Amplifiers | 11-39 |
| 11-3.36 | Application of booster amplifiers | 11-39 |
| 11-3.37 | Nomenclature | 11-39 |
| 11-3.38 | INDUCTION POTENTIOMETERS | 11-41 |
| 11-3.39 | Construction | 11-41 |
| 11-3.40 | Characteristics | 11-41 |
| 11-3.41 | TOROID-WOUND ROTARY TRANSFORMERS | 11-41 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 11 (cont) | |
| 11-3.42 | Construction | 11-41 |
| 11-3.43 | Principles of Operation | 11-41 |
| 11-3.44 | Electrical Characteristics | 11-43 |
| 11-3.45 | MICROSYNS | 11-44 |
| 11-3.46 | Principles of Operation | 11-44 |
| 11-3.47 | General Classification | 11-44 |
| 11-3.48 | Microsyn Torque Generators | 11-48 |
| 11-3.49 | Double-winding type | 11-45 |
| 11-3.50 | Single-winding type | 11-45 |
| 11-3.51 | Microsyn Signal Generator | 11-46 |
| 11-3.52 | Typical Performance Characteristics | 11-46 |
| 11-4 | LINEAR VARIABLE DIFFERENTIAL TRANS- FORMERS | 11-47 |
| 11-4.1 | GENERAL DESCRIPTION | 11-47 |
| 11-4.2 | DESIGN CHARACTERISTICS | 11-48 |
| 11-4.3 | Sensitivity Rating | 11-48 |
| 11-4.4 | Input and Output Characteristics | 11-49 |
| 11-4.5 | Phase Angle | 11-49 |
| 11-4.6 | Output Impedance | 11-49 |
| 11-4.7 | Loading | 11-49 |
| 11-4.8 | Construction | 11-50 |
| 11-4.9 | APPLICATION | 11-50 |
| 11-5 | TACHOMETER GENERATORS | 11-51 |
| 11-5.1 | GENERAL DESCRIPTION | 11-51 |
| 11-5.2 | DRAG-CUP A-C TACHOMETER GENERATOR | 11-51 |
| 11-5.3 | Theory of Operation | 11-51 |
| 11-5.4 | Equivalent Circuits and General Equations | 11-52 |
| 11-5.5 | General equation for circuit having a voltage source | 11-52 |
| 11-5.6 | General equation for a circuit having a circuit source | 11-54 |
| 11-5.7 | Speed Error | 11-54 |
| 11-5.8 | Calibration for speed errors | 11-54 |
| 11-5.9 | Voltage Errors | 11-55 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 11 (cont) | |
| 11-5.10 | Frequency Errors | 11-55 |
| 11-5.11 | Harmonics | 11-55 |
| 11-5.12 | Temperature Error and Its Compensation | 11-55 |
| 11-5.13 | Residual (or Zero-Speed) Error | 11-56 |
| 11-5.14 | Wobble | 11-56 |
| 11-5.15 | Acceleration Error | 11-57 |
| 11-5.16 | Influence of parameters on acceleration error | 11-57 |
| 11-5.17 | Commercial Units | 11-57 |
| 11-5.18 | Selection | 11-58 |
| 11-5.19 | D-C TACHOMETER GENERATORS | 11-58 |
| 11-5.20 | Deficiencies in a Homopolar Unit | 11-58 |
| 11-5.21 | Deficiencies in a Conventional-Wound Unit | 11-60 |
| 11-6 | GYROSCOPES | 11-60 |
| 11-6.1 | INTRODUCTION | 11-60 |
| 11-6.2 | DESCRIPTION AND BASIC THEORY | 11-61 |
| 11-6.3 | Description | 11-61 |
| 11-6.4 | Basic Principle of Operation | 11-61 |
| 11-6.5 | GYRO UNIT PERFORMANCE | 11-66 |
| 11-6.6 | Component Performance | 11-66 |
| 11-6.7 | Drift Rate | 11-66 |
| 11-6.8 | BASIC TYPES OF GYRO UNITS | 11-68 |
| 11-6.9 | TYPICAL APPLICATIONS | 11-68 |
| 11-6.10 | TWO-DEGREE-OF-FREEDOM GYRO UNITS | 11-68 |
| 11-6.11 | Typical Applications | 11-69 |
| 11-6.12 | Geometrical stabilization | 11-69 |
| 11-6.13 | Space integration | 11-69 |
| 11-6.14 | Angular-displacement measuring device | 11-69 |
| 11-6.15 | Departures from Ideal Performance | 11-69 |
| 11-6.16 | Nutation | 11-69 |
| 11-6.17 | SINGLE-DEGREE-OF-FREEDOM GYRO UNITS | 11-70 |
| 11-6.18 | SINGLE-AXIS RATE GYRO UNIT | 11-70 |
| 11-6.19 | Departures from Ideal Performance | 11-70 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 11 (cont) | |
| 11-6.20 | Design Considerations | 11-71 |
| 11-6.21 | SINGLE-AXIS INTEGRATING GYRO UNIT | 11-71 |
| 11-6.22 | Operating Arrangement | 11-71 |
| 11-6.23 | Modes of operation | 11-71 |
| 11-6.24 | Ideal performance | 11-73 |
| 11-6.25 | Departures from ideal performance | 11-73 |
| 11-6.26 | APPLICATION FACTORS | 11-73 |
| 11-6.27 | Electrical Power Requirements | 11-73 |
| 11-6.28 | Special Operational Problems | 11-74 |
| 11-6.29 | Impedance Levels | 11-75 |
| 11-6.30 | Availability of Design Information | 11-75 |
| 11-6.31 | INERTIAL-GUIDANCE APPLICATIONS OF GYRO UNITS | 11-75 |
| 11-6.32 | INDICATION OF THE TRUE VERTICAL | 11-76 |
| 11-7 | ANALOG-TO-DIGITAL CONVERTERS | 11-81 |
| 11-7.1 | INTRODUCTION | 11-81 |
| 11-7.2 | NUMERICAL REPRESENTATION | 11-81 |
| 11-7.3 | Binary Numbers | 11-81 |
| 11-7.4 | Binary Coded Decimal Numbers | 11-81 |
| 11-7.5 | CODING DISCS | 11-83 |
| 11-7.6 | Techniques for Avoiding Misalignment Errors | 11-83 |
| 11-7.7 | Use of two brushes | 11-83 |
| 11-7.8 | Cyclic code | 11-83 |
| 11-7.9 | Cyclic coding disc | 11-85 |
| 11-7.10 | VOLTAGE-TO-DIGITAL CODERS | 11-86 |
| 11-7.11 | Voltage Coder Operating on Time-Interval Principle | 11-86 |
| 11-7.12 | Voltage Coder Operating on Voltage-Comparison Principle | 11-88 |
| 11-8 | OTHER FORMS OF SENSING ELEMENTS | 11-88 |
| 11-8.1 | INTRODUCTION | 11-88 |
| 11-8.2 | DIFFERENTIALS | 11-88 |
| 11-8.3 | HYDRAULIC- OR MECHANICAL-AMPLIFIER IN- PUTS | 11-89 |
| 11-8.4 | PRESSURE-SENSITIVE DEVICES | 11-89 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 12 | |
| | SIGNAL CONVERTERS | |
| 12-1 | INTRODUCTION | 12-1 |
| 12-1.1 | TYPES | 12-1 |
| 12-1.2 | Modulators | 12-1 |
| 12-1.3 | Demodulators | 12-1 |
| 12-1.4 | Form of Modulation | 12-2 |
| 12-2 | MODULATORS | 12-2 |
| 12-2.1 | CHOPPER MODULATORS | 12-2 |
| 12-2.2 | Description | 12-2 |
| 12-2.3 | Characteristics | 12-3 |
| 12-2.4 | Phase shift ϕ | 12-3 |
| 12-2.5 | Drive-voltage frequency | 12-3 |
| 12-2.6 | Drive voltage | 12-3 |
| 12-2.7 | Contact rating | 12-4 |
| 12-2.8 | Chopper noise | 12-4 |
| 12-2.9 | Temperature effects | 12-4 |
| 12-2.10 | Packaging | 12-4 |
| 12-2.11 | Practical Circuits | 12-4 |
| 12-2.12 | MAGNETIC MODULATORS | 12-4 |
| 12-2.13 | Principle of Operation | 12-4 |
| 12-2.14 | Operating Circuits | 12-7 |
| 12-2.15 | Sensitivity | 12-7 |
| 12-2.16 | Harmonic Attenuation | 12-7 |
| 12-2.17 | Performance and Application of Magnetic Modulators | 12-7 |
| 12-2.18 | Core Material | 12-7 |
| 12-2.19 | ELECTRONIC MODULATORS | 12-7 |
| 12-2.20 | Operation | 12-7 |
| 12-2.21 | Bridge modulator | 12-10 |
| 12-2.22 | Typical curve | 12-10 |
| 12-2.23 | Stability | 12-10 |
| 12-3 | ELECTRONIC DEMODULATORS | 12-11 |
| 12-3.1 | DIODE DEMODULATORS | 12-11 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|------------------------------------|-------------|
| | CHAPTER 12 (cont) | |
| 12-3.2 | Operation | 12-11 |
| 12-3.3 | Ripple | 12-11 |
| 12-3.4 | Unbalance | 12-12 |
| 12-3.5 | Transfer function | 12-12 |
| 12-3.6 | Output stability | 12-12 |
| 12-3.7 | Full-Wave Demodulator | 12-12 |
| 12-3.8 | TRIODE DEMODULATORS | 12-12 |
| 12-3.9 | Keyed Demodulator | 12-12 |
| 12-3.10 | Operation | 12-12 |
| 12-3.11 | Transfer function | 12-14 |
| 12-3.12 | CHOPPER DEMODULATORS | 12-14 |
| 12-3.13 | Operation | 12-14 |
| 12-3.14 | Output filtering | 12-15 |
| 12-3.15 | Dual Use | 12-15 |
| 12-4 | DIGITAL-TO-ANALOG CONVERSION | 12-16 |
| 12-4.1 | GENERAL | 12-16 |
| 12-4.2 | Networks | 12-16 |
| 12-4.3 | Operation | 12-16 |
| 12-4.4 | Accuracy | 12-17 |
| 12-4.5 | Servomechanism | 12-18 |

TABLE OF CONTENTS

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 13 | |
| | AMPLIFIERS USED IN CONTROLLERS | |
| 13-1 | ELECTRONIC AMPLIFIERS | 13-1 |
| 13-1.1 | VACUUM TUBES | 13-1 |
| 13-1.2 | Diodes | 13-1 |
| 13-1.3 | Control of electron flow | 13-2 |
| 13-1.4 | Diodes as rectifiers | 13-2 |
| 13-1.5 | Triodes | 13-3 |
| 13-1.6 | Plate characteristics | 13-3 |
| 13-1.7 | Graphical analysis | 13-3 |
| 13-1.8 | Linear approximations | 13-4 |
| 13-1.9 | Region of operation | 13-4 |
| 13-1.10 | Linear equivalent circuits | 13-4 |
| 13-1.11 | Alternate linear equivalent circuit | 13-5 |
| 13-1.12 | Quiescent operating point | 13-5 |
| 13-1.13 | Phase shift | 13-5 |
| 13-1.14 | Pentodes and Beam-Power Tubes | 13-5 |
| 13-1.15 | Plate characteristics | 13-6 |
| 13-1.16 | Linear equivalent circuits | 13-6 |
| 13-1.17 | Graphical analysis | 13-6 |
| 13-1.18 | Interelectrode Capacitance | 13-6 |
| 13-1.19 | Tube Specifications | 13-7 |
| 13-1.20 | LINEAR ANALYSIS OF SINGLE-STAGE VACUUM-TUBE VOLTAGE AMPLIFIERS | 13-7 |
| 13-1.21 | Characteristics of Tubes Used | 13-7 |
| 13-1.22 | Simple Amplifier | 13-7 |
| 13-1.23 | Series Tube Triode Amplifier | 13-8 |
| 13-1.24 | Cascode Amplifier | 13-8 |
| 13-1.25 | Cathode Followers | 13-9 |
| 13-1.26 | Simple Feedback Amplifier | 13-11 |
| 13-1.27 | Differential Amplifiers | 13-12 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 13 (cont) | |
| 13-1.28 | Use of Pentodes | 13-13 |
| 13-1.29 | POWER AMPLIFIERS | 13-13 |
| 13-1.30 | Tubes Used in Power Amplifiers | 13-13 |
| 13-1.31 | Push-Pull Power Amplifiers | 13-14 |
| 13-1.32 | Analysis of push-pull power amplifiers | 13-15 |
| 13-1.33 | Push-pull amplifier with a-c supply | 13-15 |
| 13-1.34 | Efficiency | 13-15 |
| 13-1.35 | CASCADING AMPLIFIER STAGES | 13-15 |
| 13-1.36 | Direct-Coupled Amplifiers | 13-15 |
| 13-1.37 | Problems encountered in direct-coupled amplifiers | 13-18 |
| 13-1.38 | Drift-compensated direct-coupled amplifier | 13-19 |
| 13-1.39 | Bridge circuits | 13-19 |
| 13-1.40 | A-C Coupled Amplifiers | 13-21 |
| 13-1.41 | Two-stage a-c coupled amplifier | 13-21 |
| 13-1.42 | FEEDBACK AMPLIFIERS | 13-23 |
| 13-1.43 | Advantages | 13-23 |
| 13-1.44 | Disadvantages | 13-25 |
| 13-1.45 | PROBLEMS ENCOUNTERED IN USE OF ELECTRONIC AMPLIFIERS AS SERVO COMPONENTS .. | 13-26 |
| 13-1.46 | Reliability | 13-26 |
| 13-1.47 | Construction | 13-27 |
| 13-1.48 | Maintenance | 13-27 |
| 13-1.49 | Quadrature Signals | 13-27 |
| 13-1.50 | Complete Amplifier | 13-27 |
| 13-1.51 | Details of a Typical Servo Amplifier | 13-28 |
| 13-1.52 | THYRATRON AMPLIFIERS | 13-28 |
| 13-1.53 | Description of Thyatron | 13-28 |
| 13-1.54 | Thyatron Characteristics | 13-30 |
| 13-1.55 | Thyatron Amplifier with Resistive Load | 13-30 |
| 13-1.56 | Control of Load Voltage | 13-31 |
| 13-1.57 | Thyatron-Amplifier Loads | 13-32 |
| 13-1.58 | Resistive loads | 13-33 |
| 13-1.59 | Inductive loads | 13-33 |

TABLE OF CONTENTS

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 13 (cont) | |
| 13-1.60 | Battery, capacitive, and separately excited d-c motor loads | 13-34 |
| 13-1.61 | Dynamic Performance | 13-35 |
| 13-1.62 | Exception to time-constant rule | 13-35 |
| 13-1.63 | D-C POWER SUPPLIES FOR ELECTRONIC AMPLIFIERS | 13-35 |
| 13-1.64 | Types of Rectifiers | 13-35 |
| 13-1.65 | Power-Supply Circuits | 13-35 |
| 13-1.66 | Design of D-C Power Supplies | 13-36 |
| 13-1.67 | Typical electronic regulator | 13-37 |
| 13-2 | TRANSISTOR AMPLIFIERS | 13-38 |
| 13-2.1 | BASIC PRINCIPLES | 13-38 |
| 13-2.2 | Operating Characteristics of Temperature-Limited Vacuum Diode | 13-38 |
| 13-2.3 | Transistor Operation | 13-39 |
| 13-2.4 | Advantages and disadvantages of transistors | 13-39 |
| 13-2.5 | High-frequency operation | 13-39 |
| 13-2.6 | Medium-frequency operation | 13-39 |
| 13-2.7 | High-power applications | 13-40 |
| 13-2.8 | Switching applications | 13-40 |
| 13-2.9 | Summary | 13-40 |
| 13-2.10 | BASIC THEORY OF JUNCTION DIODES AND TRANSISTORS | 13-40 |
| 13-2.11 | Electron Current | 13-40 |
| 13-2.12 | Hole Current | 13-40 |
| 13-2.13 | Material Types | 13-41 |
| 13-2.14 | Junctions | 13-41 |
| 13-2.15 | Junction Diodes | 13-41 |
| 13-2.16 | Junction Transistors | 13-42 |
| 13-2.17 | Special transistor types | 13-42 |
| 13-2.18 | Characteristics of transistor materials | 13-42 |
| 13-2.19 | ANALYSIS OF TRANSISTOR CHARACTERISTICS | 13-43 |
| 13-2.20 | Transistor Model | 13-43 |
| 13-2.21 | Piecewise linear model | 13-44 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 13 (cont) | |
| 13-2.22 | Incremental model | 13-44 |
| 13-2.23 | Hybrid Parameters | 13-46 |
| 13-2.24 | Frequency Dependence | 13-46 |
| 13-2.25 | Temperature Sensitivity | 13-47 |
| 13-2.26 | Nonuniformity | 13-47 |
| 13-2.27 | Noise Factor | 13-50 |
| 13-2.28 | Microphonics and Vibration Effects | 13-50 |
| 13-2.29 | Maximum Collector Voltage | 13-50 |
| 13-2.30 | Maximum Power Dissipation | 13-50 |
| 13-2.31 | TRANSISTOR AMPLIFIER CIRCUITS | 13-50 |
| 13-2.32 | Grounded-Emitter Amplifier | 13-50 |
| 13-2.33 | Grounded-Base Amplifier | 13-50 |
| 13-2.34 | Grounded-Collector Amplifier | 13-51 |
| 13-2.35 | Phase-Inverter and Difference-Amplifier Circuits | 13-51 |
| 13-2.36 | High Power Amplifiers | 13-51 |
| 13-2.37 | Maximum power output | 13-52 |
| 13-2.38 | Biasing | 13-52 |
| 13-2.39 | Typical Two-Stage Transistor Amplifier | 13-54 |
| 13-2.40 | Direct-Coupled Amplifiers | 13-54 |
| 13-2.41 | State of the Art | 13-55 |
| 13-2.42 | A-C POWER AMPLIFIER DESIGN | 13-55 |
| 13-3 | MAGNETIC AMPLIFIERS | 13-58 |
| 13-3.1 | BASIC CONSIDERATIONS | 13-58 |
| 13-3.2 | Functions of Magnetic Amplifiers in Servo Systems .. | 13-58 |
| 13-3.3 | Features of Magnetic Amplifiers | 13-58 |
| 13-3.4 | Application Problems of Magnetic Amplifiers | 13-58 |
| 13-3.5 | Temperature Limitations | 13-58 |
| 13-3.6 | Design Difficulties | 13-58 |
| 13-3.7 | PRINCIPLES OF OPERATION | 13-59 |
| 13-3.8 | Single-Core, Single-Rectifier Circuit | 13-59 |
| 13-3.9 | Operating cycle of a single-core circuit | 13-59 |
| 13-3.10 | Exciting and conducting periods | 13-60 |
| 13-3.11 | Reset period | 13-60 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 13 (cont) | |
| 13-3.12 | Control limits | 13-60 |
| 13-3.13 | Control characteristics | 13-61 |
| 13-3.14 | Analysis limitations | 13-61 |
| 13-3.15 | Analysis extension | 13-62 |
| 13-3.16 | TYPICAL CIRCUITS | 13-62 |
| 13-3.17 | Half-Cycle (Ramey) Circuit | 13-62 |
| 13-3.18 | Single-Ended, Two-Core Circuits | 13-62 |
| 13-3.19 | Operation of two-core circuits | 13-63 |
| 13-3.20 | Reversible-Polarity and Reversible-Phase Circuits | 13-63 |
| 13-3.21 | Example of reversible-phase amplifier | 13-63 |
| 13-3.22 | ANALYTICAL REPRESENTATION OF MAGNETIC AMPLIFIERS | 13-64 |
| 13-3.23 | Dynamic Performance | 13-64 |
| 13-3.24 | Dynamic response | 13-65 |
| 13-3.25 | Accuracy of prediction | 13-65 |
| 13-3.26 | Analytical Representation | 13-65 |
| 13-3.27 | Effective control-circuit resistance | 13-65 |
| 13-3.28 | Limitations in analytical representation | 13-66 |
| 13-3.29 | Conclusions | 13-66 |
| 13-3.30 | PERFORMANCE OF MAGNETIC AMPLIFIERS | 13-66 |
| 13-3.31 | Ranges of Input and Output Power | 13-66 |
| 13-3.32 | Control Characteristics | 13-66 |
| 13-3.33 | Figure of Merit | 13-66 |
| 13-3.34 | Typical values of Figure of Merit | 13-68 |
| 13-3.35 | Expression for Figure of Merit | 13-69 |
| 13-3.36 | Largest factor | 13-70 |
| 13-3.37 | Effects of dimensions | 13-70 |
| 13-3.38 | Leakage effects | 13-70 |
| 13-3.39 | Temperature effects | 13-70 |
| 13-3.40 | CONSTRUCTION OF MAGNETIC AMPLIFIERS | 13-70 |
| 13-3.41 | Methods of Core Construction | 13-70 |
| 13-3.42 | Relative merits | 13-71 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 13 (cont) | |
| 13-3.43 | Core Materials | 13-71 |
| 13-3.44 | Applications | 13-71 |
| 13-3.45 | Rectifiers | 13-72 |
| 13-3.46 | Characteristics of selenium rectifiers | 13-72 |
| 13-3.47 | Characteristics of germanium rectifiers | 13-72 |
| 13-3.48 | Characteristics of silicon rectifiers | 13-72 |
| 13-3.49 | Factors governing choice of rectifier type | 13-72 |
| 13-3.50 | SPECIFICATIONS AND DESIGN | 13-72 |
| 13-3.51 | Specifications | 13-72 |
| 13-3.52 | Approach to Design | 13-73 |
| 13-4 | ROTARY ELECTRIC AMPLIFIERS | 13-73 |
| 13-4.1 | TYPES OF ROTARY ELECTRIC AMPLIFIERS | 13-73 |
| 13-4.2 | Basic Principles | 13-73 |
| 13-4.3 | Power-Handling and Time-Constant Characteristics | 13-73 |
| 13-4.4 | Basic Features | 13-74 |
| 13-4.5 | Types of Rotary Electric Amplifiers | 13-74 |
| 13-4.6 | Excitation | 13-74 |
| 13-4.7 | Single-shunt-winding machine | 13-76 |
| 13-4.8 | Armature reaction | 13-76 |
| 13-4.9 | Cross-field machine | 13-76 |
| 13-4.10 | Other multistage machines | 13-76 |
| 13-4.11 | CHARACTERISTICS OF ROTARY ELECTRIC AMPLIFIERS | 13-77 |
| 13-4.12 | Steady-State and Transient Characteristics | 13-77 |
| 13-4.13 | Deriving input-output characteristics | 13-77 |
| 13-4.14 | Simplifying characteristic derivation | 13-77 |
| 13-4.15 | Effects of saturation and variable speed | 13-78 |
| 13-4.16 | Results of saturation | 13-79 |
| 13-4.17 | Results of speed variations | 13-79 |
| 13-4.18 | Linear operation | 13-79 |
| 13-4.19 | Derivation of dynamic characteristics | 13-79 |
| 13-4.20 | PARAMETERS OF D-C ROTARY AMPLIFIERS | 13-80 |
| 13-4.21 | Fundamental Requirements | 13-80 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 13 (cont) | |
| 13-4.22 | Stored energy | 13-82 |
| 13-4.23 | Resistance of field winding | 13-82 |
| 13-4.24 | Dissipated power | 13-82 |
| 13-4.25 | Basic information for field-system design | 13-82 |
| 13-4.26 | Power amplification | 13-83 |
| 13-4.27 | PROBLEMS ENCOUNTERED IN THE USE OF ROTARY ELECTRIC AMPLIFIERS IN SERVO APPLICATIONS | 13-83 |
| 13-4.28 | Design Factors | 13-83 |
| 13-4.29 | SELECTION OF ROTARY ELECTRIC AMPLIFIERS FOR CONTROL PURPOSES | 13-85 |
| 13-4.30 | Controlling Factors | 13-85 |
| 13-4.31 | TYPICAL CHARACTERISTICS AND DESIGN DATA OF SOME ROTARY ELECTRIC AMPLIFIERS | 13-86 |
| 13-4.32 | Information Available from Manufacturers | 13-86 |
| 13-4.33 | Dynamic Performance | 13-87 |
| 13-5 | RELAY AMPLIFIERS | 13-87 |
| 13-5.1 | DEFINITION | 13-87 |
| 13-5.2 | ADVANTAGES AND DISADVANTAGES | 13-87 |
| 13-5.3 | RELAY CHARACTERISTICS | 13-88 |
| 13-5.4 | Description of Operation | 13-88 |
| 13-5.5 | Usage of Relays | 13-88 |
| 13-5.6 | SINGLE-SIDED RELAY AMPLIFIERS FOR SPEED CONTROL | 13-88 |
| 13-5.7 | Typical Amplifiers | 13-88 |
| 13-5.8 | Operation | 13-89 |
| 13-5.9 | REVERSIBLE MOTOR-SHAFT ROTATION | 13-90 |
| 13-5.10 | Servomechanism Applications | 13-90 |
| 13-5.11 | Relay-Amplifier Operation for Positional Control | 13-90 |
| 13-5.12 | Basic operation | 13-90 |
| 13-5.13 | Phase-Sensitive Relay Amplifiers | 13-90 |
| 13-5.14 | Typical types | 13-90 |
| 13-5.15 | Operation of single-stage phase-sensitive relay amplifier | 13-90 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 13 (cont) | <i>Page</i> |
|------------------|---|-------------|
| 13-5.16 | Choice of circuit components | 13-91 |
| 13-5.17 | POLARIZED RELAYS | 13-93 |
| 13-5.18 | Purpose | 13-93 |
| 13-5.19 | Advantages and Disadvantages | 13-98 |
| 13-5.20 | STATIC CHARACTERISTICS OF RELAYS | 13-98 |
| 13-5.21 | Idealized Relay | 13-98 |
| 13-5.22 | Sensitivity | 13-98 |
| 13-5.23 | Contact Rating | 13-99 |
| 13-5.24 | DYNAMIC CHARACTERISTICS OF RELAYS | 13-100 |
| 13-5.25 | Response Time | 13-100 |
| 13-5.26 | Nature of time delay | 13-100 |
| 13-5.27 | General Considerations | 13-100 |
| 13-5.28 | PARAMETER MEASUREMENT | 13-101 |
| 13-5.29 | Relay Operating Time | 13-101 |
| 13-5.30 | PROBLEMS ENCOUNTERED WITH RELAY AMPLIFIERS | 13-102 |
| 13-5.31 | R/δ , A Figure of Merit | 13- 02 |
| 13-5.32 | Relay Life | 13-102 |
| 13-5.33 | False Operation | 13-103 |
| 13-6 | HYDRAULIC AMPLIFIERS | 13-105 |
| 13-6.1 | INTRODUCTION | 13-105 |
| 13-6.2 | Description and Usage | 13-105 |
| 13-6.3 | Characteristics | 13-109 |
| 13-6.4 | TRANSLATIONAL HYDRAULIC AMPLIFIERS | 13-110 |
| 13-6.5 | Spool-Valve Type | 13-110 |
| 13-6.6 | Equivalent source representation of four-way spool valve | 13-110 |
| 13-6.7 | Flow equations of underlapped four-way spool valve | 13-110 |
| 13-6.8 | Dimensionalizing flow-gain and conductance plots of underlapped four-way spool valve | 13-119 |
| 13-6.9 | Small-signal gain and conductance parameters of underlapped four-way spool valve | 13-119 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 13 (cont) | |
| 13-6.10 | Balanced-load steady-state characteristics of underlapped four-way spool valve | 13-121 |
| 13-6.11 | Flow equations of four-way spool valve with radial clearance | 13-122 |
| 13-6.12 | Dimensionalizing flow-gain and conductance plots of four-way spool valve with radial clearance | 13-124 |
| 13-6.13 | Plate-Valve Type | 13-128 |
| 13-6.14 | Double Nozzle-Baffle-Valve Type | 13-129 |
| 13-6.15 | Flow equations of double nozzle-baffle valve | 13-130 |
| 13-6.16 | Double nozzle-baffle amplifier with balanced load | 13-131 |
| 13-6.17 | Dimensionalizing pressure-gain and resistance plots of double nozzle-baffle amplifier | 13-132 |
| 13-6.18 | Pressure-Control and Flow-Control Valve Amplifiers | 13-134 |
| 13-6.19 | ROTARY HYDRAULIC AMPLIFIERS | 13-136 |
| 13-6.20 | Characteristics | 13-136 |
| 13-6.21 | INTERACTION OF LOAD AND VALVE | 13-137 |
| 13-6.22 | Equivalent Hydraulic Circuit of Amplifier | 13-138 |
| 13-6.23 | Block Diagram | 13-148 |
| 13-6.24 | DYNAMIC RESPONSE OF ROTARY PUMP AND LOAD | 13-149 |
| 13-6.25 | DYNAMIC RESPONSE OF TRANSLATIONAL AMPLIFIER AND LOAD | 13-150 |
| 13-6.26 | PROBLEMS ENCOUNTERED IN USE OF HYDRAULIC AMPLIFIERS | 13-150 |
| 13-6.27 | HYDRAULIC-CIRCUIT ELEMENTS | 13-155 |
| 13-6.28 | ILLUSTRATIVE EXAMPLE | 13-158 |
| 13-6.29 | List of Pertinent Equations and Parameters | 13-158 |
| 13-6.30 | Calculation of Valve Constants | 13-158 |
| 13-6.31 | Calculation of Other Constants | 13-159 |
| 13-6.32 | Calculation of Coefficients a and b | 13-159 |
| 13-6.33 | Results, Simplification, and Significance | 13-160 |
| 13-7 | PNEUMATIC AMPLIFIERS | 13-160 |
| 13-7.1 | INTRODUCTION | 13-162 |
| 13-7.2 | PNEUMATIC VALVES | 13-163 |

TABLE OF CONTENTS (cont)

Paragraph

CHAPTER 13 (cont)

| | | |
|---------|---|--------|
| 13-7.3 | STATIC CHARACTERISTICS OF PNEUMATIC VALVES | 13-163 |
| 13-7.4 | Orifice Flow | 13-163 |
| 13-7.5 | Nondimensional Flow | 13-165 |
| 13-7.6 | Equivalent Source | 13-168 |
| 13-7.7 | Plate-Valve Static Characteristics | 13-169 |
| 13-7.8 | Conical-Plug Valve Static Characteristics | 13-170 |
| 13-7.9 | Nozzle-Baffle Valve Static Characteristics | 13-170 |
| 13-7.10 | DYNAMIC BEHAVIOR OF PNEUMATIC AMPLIFIERS | 13-173 |
| 13-7.11 | Equivalent-Circuit Elements for Pneumatic Systems | 13-173 |
| 13-7.12 | Resistance | 13-173 |
| 13-7.13 | Capacitance | 13-175 |
| 13-7.14 | Inertance | 13-175 |
| 13-7.15 | Time constant | 13-176 |
| 13-7.16 | Pneumatic Equivalents for Dashpots, Springs, and Masses | 13-176 |
| 13-7.17 | Dynamic Behavior of Four-Way Valves | 13-176 |
| 13-7.18 | Dynamic Behavior of Three-Way Valves | 13-177 |
| 13-7.19 | Typical Performance of Low-Pressure Three-Way Valves | 13-178 |
| 13-7.20 | ADVANTAGES AND DISADVANTAGES OF PNEUMATIC SYSTEMS | 13-180 |
| 13-7.21 | Advantages | 13-180 |
| 13-7.22 | Disadvantages | 13-180 |
| 13-8 | MECHANICAL AMPLIFIERS | 13-181 |
| 13-8.1 | BASIC TYPES | 13-181 |
| 13-8.2 | Variable-Speed-Output Mechanical Amplifier | 13-181 |
| 13-8.3 | Ball-disc integrator | 13-182 |
| 13-8.4 | Cone-and-disc-amplifier | 13-182 |
| 13-8.5 | Variable-Torque-Output Mechanical Amplifier | 13-182 |
| 13-8.6 | STATIC CHARACTERISTICS OF MECHANICAL AMPLIFIERS | 13-184 |
| 13-8.7 | Ideal and Actual Characteristics | 13-184 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 13 (cont) | |
| 13-8.8 | Static Characteristics of Integrator Type Mechanical Amplifiers | 13-185 |
| 13-8.9 | Static Characteristics of Capstan Type Amplifier | 13-187 |
| 13-8.10 | DYNAMIC BEHAVIOR OF MECHANICAL AMPLIFIERS | 13-187 |
| 13-8.11 | Capstan Amplifier | 13-187 |
| 13-8.12 | Integrator Amplifier | 13-191 |
| 13-8.13 | OTHER MECHANICAL AMPLIFIERS | 13-191 |
| 13-8.14 | Clutch-Type Amplifiers | 13-191 |
| 13-8.15 | PROBLEMS ENCOUNTERED WITH MECHANICAL AMPLIFIERS | 13-195 |
| 13-8.16 | Capstan Amplifiers | 13-195 |
| 13-8.17 | Integrator Amplifiers | 13-195 |
| 13-8.18 | Clutch Amplifiers | 13-196 |

TABLE OF CONTENTS

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--|-------------|
| | CHAPTER 14 | |
| | POWER ELEMENTS USED IN CONTROLLERS | |
| 14-1 | INTRODUCTION | 14-1 |
| 14-2 | DIRECT-CURRENT MOTORS | 14-1 |
| 14-2.1 | USAGE OF D-C MOTORS | 14-1 |
| 14-2.2 | STATIC CHARACTERISTICS OF D-C MOTORS | 14-7 |
| 14-2.3 | Armature Control with Constant Field Current | 14-7 |
| 14-2.4 | Field Control with Constant Armature Current | 14-10 |
| 14-2.5 | Field Control with Constant Armature Voltage | 14-10 |
| 14-2.6 | Series Motor Control | 14-10 |
| 14-2.7 | DYNAMIC CHARACTERISTICS OF D-C MOTORS | 14-11 |
| 14-2.8 | Armature Control with Constant Field Current | 14-12 |
| 14-2.9 | Field Control with Constant Armature Current | 14-12 |
| 14-2.10 | Field Control with Constant Armature Voltage | 14-13 |
| 14-2.11 | Series Motor Control | 14-15 |
| 14-2.12 | Rotary Electric Amplifier Control | 14-16 |
| 14-2.13 | Electronic Amplifier Control | 14-17 |
| 14-2.14 | MEASUREMENT OF D-C MOTOR PARAMETERS .. | 14-17 |
| 14-2.15 | Armature Resistance | 14-18 |
| 14-2.16 | Armature Inductance | 14-18 |
| 14-2.17 | Field Resistance | 14-18 |
| 14-2.18 | Field Inductance | 14-18 |
| 14-2.19 | Motor Constant | 14-19 |
| 14-2.20 | Armature Inertia | 14-19 |
| 14-2.21 | Viscous Damping | 14-20 |
| 14-2.22 | Typical Parameter Values | 14-20 |
| 14-2.23 | D-C TORQUE MOTORS | 14-20 |
| 14-2.24 | Description of Typical Torque Motors | 14-20 |
| 14-2.25 | Static Characteristics | 14-20 |
| 14-2.26 | Dynamic Characteristics | 14-22 |
| 14-2.27 | Modified Servomotors | 14-22 |
| 14-2.28 | CONVERSION FACTORS AND UNITS | 14-23 |
| 14-3 | ALTERNATING-CURRENT MOTORS | 14-25 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 14 (cont) | <i>Page</i> |
|------------------|--|-------------|
| 14-3.1 | TYPES OF A-C MOTORS USED IN SERVOMECHANISMS | 14-25 |
| 14-3.2 | Amplifier Control of Servomotors | 14-25 |
| 14-3.3 | Types of A-C Motors Used in Relay Applications | 14-29 |
| 14-3.4 | Typical relay-servo circuits | 14-30 |
| 14-3.5 | STATIC CHARACTERISTICS OF A-C MOTORS | 14-31 |
| 14-3.6 | Typical 2-Phase Servomotor | 14-31 |
| 14-3.7 | Relay Servos | 14-31 |
| 14-3.8 | DYNAMIC BEHAVIOR OF 2-PHASE SERVOMOTORS | 14-31 |
| 14-3.9 | Equations | 14-31 |
| 14-3.10 | Alternatives | 14-33 |
| 14-3.11 | Torque at partial excitation | 14-34 |
| 14-3.12 | Nondimensional damping | 14-34 |
| 14-3.13 | Motor equations | 14-35 |
| 14-3.14 | FIGURE OF MERIT | 14-37 |
| 14-3.15 | COMPARISON OF TYPICAL MOTORS | 14-37 |
| 14-4 | HYDRAULIC MOTORS | 14-37 |
| 14-4.1 | INTRODUCTION | 14-37 |
| 14-4.2 | STATIC CHARACTERISTICS OF PISTON-TYPE ROTARY MOTORS | 14-37 |
| 14-4.3 | STATIC EQUATIONS OF TRANSLATORY MOTOR (MOVING PISTON) | 14-39 |
| 14-4.4 | DYNAMIC BEHAVIOR OF HYDRAULIC TRANSMISSIONS | 14-39 |
| 14-4.5 | APPROXIMATE DYNAMIC BEHAVIOR OF HYDRAULIC TRANSMISSION | 14-43 |
| 14-4.6 | PARAMETER EVALUATION FOR ROTARY MOTOR | 14-43 |
| 14-4.7 | PROBLEMS ENCOUNTERED WITH HYDRAULIC MOTORS | 14-44 |
| 14.5 | PNEUMATIC MOTORS | 14-44 |
| 14-5.1 | PRINCIPAL TYPES | 14-44 |
| 14-5.2 | STATIC BEHAVIOR | 14-45 |
| 14-5.3 | DYNAMIC BEHAVIOR | 14-45 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|---|--|-------------|
| CHAPTER 14 (cont) | | |
| 14-5.4 | DIFFICULTIES ENCOUNTERED WITH PNEU- MATIC MOTORS | 14-46 |
| 14.6 | MAGNETIC-PARTICLE CLUTCHES | 14-47 |
| 14-6.1 | DESCRIPTION | 14-47 |
| 14-6.2 | METHODS OF USE | 14-47 |
| 14-6.3 | ADVANTAGES | 14-47 |
| 14-6.4 | DISADVANTAGES | 14-47 |
| 14-6.5 | STATIC BEHAVIOR | 14-50 |
| 14-6.6 | DYNAMIC BEHAVIOR | 14-51 |
| 14-6.7 | Voltage-to-Position Relationship | 14-51 |
| 14-6.8 | LIFE EXPECTANCY | 14-51 |
| CHAPTER 15 | | |
| MECHANICAL AUXILIARIES USED IN CONTROLLERS | | |
| 15-1 | GEAR TRAINS | 15-1 |
| 15-1.1 | PURPOSE | 15-1 |
| 15-1.2 | DEFINITIONS | 15-1 |
| 15-1.3 | GEAR TYPES | 15-3 |
| 15-1.4 | DESIGN FUNDAMENTALS | 15-5 |
| 15-1.5 | Backlash | 15-5 |
| 15-1.6 | Dynamic Load | 15-5 |
| 15-1.7 | Gear Accuracy | 15-7 |
| 15-1.8 | Beam Strength of Teeth | 15-8 |
| 15-1.9 | Limit Load for Wear | 15-8 |
| 15-1.10 | NONIDEAL CHARACTERISTICS OF GEARS AND GEAR TRAINS | 15-12 |
| 15-1.11 | Inaccuracies | 15-12 |
| 15-1.12 | Friction | 15-13 |
| 15-1.13 | Inertia | 15-13 |
| 15-1.14 | Backlash | 15-13 |
| 15-1.15 | Compliance | 15-15 |
| 15-2 | MECHANICAL DIFFERENTIALS | 15-16 |
| 15-2.1 | PURPOSE | 15-16 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 15 (cont) | <i>Page</i> |
|------------------|-------------------------------------|-------------|
| 15-2.2 | GEARED DIFFERENTIALS | 15-16 |
| 15-2.3 | DIFFERENTIAL LINKAGES | 15-17 |
| 15-3 | LINKAGES AND LEVERS | 15-17 |
| 15-3.1 | BASIC PURPOSE | 15-17 |
| 15-3.2 | PRACTICAL APPLICATIONS | 15-17 |
| 15-3.3 | EXAMPLES | 15-19 |
| 15-3.4 | NONIDEAL CHARACTERISTICS | 15-19 |
| 15-3.5 | Mass of Links and Cranks | 15-19 |
| 15-3.6 | Compliance | 15-20 |
| 15-3.7 | Backlash in Pivots and Slides | 15-20 |
| 15-4 | SHEAVES AND TAPES | 15-20 |
| 15-4.1 | PURPOSE | 15-20 |
| 15-4.2 | STRESS | 15-20 |
| 15-4.3 | TENSION | 15-21 |
| 15-4.4 | SHEAVE SIZES | 15-21 |
| 15-4.5 | COMPLIANCE | 15-21 |
| 15-5 | MECHANICAL COUPLING DEVICES | 15-22 |
| 15-5.1 | COUPLINGS | 15-22 |
| 15-5.2 | KEYS AND SPLINES | 15-25 |
| 15-5.3 | Keys | 15-25 |
| 15-5.4 | Splines | 15-25 |
| 15-5.5 | Involute Splines | 15-29 |
| 15-6 | BEARINGS | 15-29 |
| 15-6.1 | ROLLER BEARINGS | 15-29 |
| 15-6.2 | BALL BEARINGS | 15-29 |
| 15-6.3 | LUBRICATION | 15-32 |
| 15-6.4 | FRICTION | 15-32 |
| 15-6.5 | SLEEVE BEARINGS | 15-33 |
| 15-6.6 | MISCELLANEOUS BEARINGS | 15-34 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 16 | <i>Page</i> |
|------------------|--|-------------|
| | TYPICAL PROCEDURE | |
| 16-1 | INTRODUCTION | 16-1 |
| 16-2 | GATHERING OF SPECIFICATIONS | 16-2 |
| 16-3 | CHOICE OF TRIAL COMPONENTS | 16-4 |
| 16-3.1 | GENERAL | 16-4 |
| 16-3.2 | DETERMINATION OF MOTOR SIZE | 16-4 |
| 16-3.3 | DETERMINATION OF GEAR RATIO | 16-5 |
| 16-3.4 | SELECTION OF OUTPUT MEMBER AND REMAIN- ING ELEMENTS | 16-6 |
| 16-4 | ANALYSIS OF TRIAL SYSTEM | 16-13 |
| 16-5 | MODIFICATION OR REDESIGN OF TRIAL SYSTEM | 16-13 |
| 16-6 | CONSTRUCTION AND TEST OF EXPERIMENTAL EQUIPMENT | 16-14 |
| 16-7 | TRANSLATION OF EXPERIMENTAL EQUIPMENT INTO PRODUCTION MODEL | 16-14 |
| 16.8 | ILLUSTRATIVE EXAMPLE — DESIGN OF A SERVO DATA REPEATER | 16-14 |
| 16-8.1 | NATURE OF THE MEASUREMENT PROBLEM | 16-14 |
| 16-8.2 | DESIGN OF THE SERVO | 16-15 |
| 16-8.3 | Accuracy Determinations | 16-15 |
| 16-8.4 | Antenna characteristics | 16-15 |
| 16-8.5 | Synchro accuracies | 16-15 |
| 16-8.6 | Dials | 16-16 |
| 16-8.7 | Friction and Drift | 16-16 |
| 16-8.8 | Dynamic errors | 16-16 |
| 16-8.9 | Choice of Servo Compensation Method | 16-16 |
| 16-8.10 | Design of the Damper-Stabilized Servo | 16-18 |
| 16-8.11 | Detailed analysis | 16-18 |
| 16-8.12 | Performance of the completed servo | 16-21 |
| 16-8.13 | Improvement by using dual-mode operation | 16-23 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | CHAPTER 17 | <i>Page</i> |
|---|---|-------------|
| | REPRESENTATIVE DESIGNS | |
| 17-1 | INTRODUCTION | 17-1 |
| 17-2 | A SERVO SYSTEM FOR A TRACKING-RADAR AN- TENNA | 17-1 |
| 17-2.1 | GENERAL | 17-1 |
| 17-2.2 | PURPOSE | 17-1 |
| 17-2.3 | OPERATION | 17-1 |
| 17-2.4 | OPERATIONAL BLOCK DIAGRAM | 17-3 |
| 17-2.5 | NOISE | 17-4 |
| 17-2.6 | DESIGN CHARACTERISTICS | 17-4 |
| 17-3 | A POWER CONTROL SYSTEM FOR THE M-38 FIRE- CONTROL SYSTEM | 17-5 |
| 17-3.1 | GENERAL | 17-5 |
| 17-3.2 | OPERATION | 17-5 |
| 17-3.3 | OPERATIONAL BLOCK DIAGRAM | 17-8 |
| 17-3.4 | DESIGN CHARACTERISTICS | 17-9 |
| CHAPTER 18 AUXILIARIES ASSOCIATED WITH SERVOMECHANISMS | | |
| 18-1 | AUXILIARY PUMPS | 18-1 |
| 18-1.1 | PURPOSE | 18-1 |
| 18-1.2 | TYPES OF AUXILIARY PUMPS | 18-1 |
| 18-1.3 | Gear Pumps | 18-1 |
| 18-1.4 | Vane Pumps | 18-2 |
| 18-1.5 | Piston Pumps | 18-3 |
| 18-1.6 | MAINTENANCE | 18-3 |
| 18-1.7 | LEAKAGE AND DRAINAGE | 18-3 |
| 18-1.8 | COST | 18-4 |
| 18-2 | HYDRAULIC AUXILIARIES | 18-4 |
| 18-2.1 | HYDRAULIC SYSTEMS INCORPORATING AUX- ILIARIES | 18-4 |
| 18-2.2 | CHECK VALVES | 18-4 |
| 18-2.3 | Ball Check Valves | 18-4 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|--------------------------------------|-------------|
| | CHAPTER 18 (cont) | |
| 18-2.4 | PRESSURE-RELIEF VALVES | 18-6 |
| 18-2.5 | PRESSURE-REGULATING VALVES | 18-6 |
| 18-2.6 | ACCUMULATORS | 18-8 |
| 18-2.7 | Gravity Accumulators | 18-8 |
| 18-2.8 | Hydropneumatic Accumulators | 18-9 |
| 18-2.9 | UNLOADING VALVES | 18-10 |
| 18-3 | ROTARY JOINTS | 18-12 |
| 18-3.1 | DYNAMIC SEALS | 18-12 |
| 18-3.2 | Glands | 18-12 |
| 18-3.3 | O-Rings | 18-12 |
| 18-3.4 | U-Cup and V-Ring Packings | 18-12 |
| 18-3.5 | Shaft Seals | 18-12 |
| 18-3.6 | Face Seals | 18-13 |
| 18-3.7 | High-Pressure Seals | 18-14 |
| 18-3.8 | Friction | 18-15 |
| 18-4 | LIMIT STOPS AND POSITIVE STOPS | 18-15 |
| 18-4.1 | PURPOSE | 18-15 |
| 18-4.2 | CHARACTERISTICS | 18-15 |
| 18-4.3 | LIMIT STOPS | 18-16 |
| 18-4.4 | POSITIVE STOPS | 18-18 |
| 18-4.5 | Buffers | 18-18 |

CHAPTER 19 CONSTRUCTIONAL TECHNIQUES

| | | |
|--------|---|------|
| 19-1 | BASIC CONSIDERATIONS | 19-1 |
| 19-2 | COMPONENT LAYOUT | 19-1 |
| 19-2.1 | PHYSICAL ARRANGEMENT | 19-1 |
| 19-2.2 | THERMAL CONSIDERATIONS AND HEAT GEN- ERATION | 19-2 |
| 19-2.3 | Radiation | 19-4 |
| 19-2.4 | Free Convection by Air | 19-4 |
| 19-2.5 | Conduction | 19-6 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|---|--|-------------|
| CHAPTER 19 (cont) | | |
| 19-2.6 | EQUIPMENT AND PERSONNEL SAFETY MEASURES | 19-6 |
| 19-3 | VIBRATION ISOLATION | 19-8 |
| 19-4 | SHOCK ISOLATION | 19-10 |
| CHAPTER 20 | | |
| SUPPLEMENTARY TABLES, FORMULAS, AND CHARTS | | |
| 20-1 | MASS MOMENT OF INERTIA | 20-1 |
| 20-1.1 | DEFINITION | 20-1 |
| 20-1.2 | DATA FOR EQUATIONS | 20-1 |
| 20-1.3 | PARALLEL-AXIS THEOREM | 20-1 |
| 20-1.4 | PRINCIPAL AXES OF INERTIA | 20-1 |
| 20-1.5 | PRODUCT OF INERTIA | 20-1 |
| 20-1.6 | INERTIA WITH RESPECT TO A LINE THROUGH THE ORIGIN | 20-4 |
| 20-1.7 | TABULATED MOMENTS OF INERTIA | 20-4 |
| 20-1.8 | COMPLICATED SHAPES | 20-4 |
| 20-2 | DAMPING AND FRICTION | 20-4 |
| 20-2.1 | VISCOSITY | 20-4 |
| 20-2.2 | Definition | 20-4 |
| 20-2.3 | Absolute (Dynamic) Viscosity | 20-4 |
| 20-2.4 | Kinematic Viscosity | 20-5 |
| 20-2.5 | Effect of Temperature | 20-8 |
| 20-2.6 | FRICTION | 20-8 |
| 20-2.7 | Coefficient of Friction | 20-8 |
| 20-2.8 | Characteristics of Coefficient of Friction | 20-9 |
| 20-3 | SPRINGS | 20-9 |
| 20-3.1 | HELICAL SPRINGS | 20-9 |
| 20-3.2 | Stress | 20-9 |
| 20-3.3 | Deflection | 20-9 |
| 20-3.4 | Torsional Elasticity | 20-9 |
| 20-3.5 | Design Table | 20-9 |

TABLE OF CONTENTS (cont)

| <i>Paragraph</i> | | <i>Page</i> |
|------------------|---|-------------|
| | CHAPTER 20 (cont) | |
| 20-3.6 | CANTILEVER SPRINGS | 20-11 |
| 20-3.7 | Definition | 20-11 |
| 20-3.8 | Stress and Deflection | 20-11 |
| 20-3.9 | Tensile Elastic Properties | 20-11 |
| 20-3.10 | TORSION BAR SPRINGS | 20-11 |
| 20-3.11 | Definition | 20-11 |
| 20-3.12 | Stress and Torque | 20-11 |
| 20-3.13 | VIBRATION IN SPRINGS | 20-14 |
| 20-3.14 | Natural Frequency | 20-14 |
| 20-4 | MISCELLANEOUS CONSTANTS OF FLUIDS AND TUBING USED IN HYDRAULIC SYSTEMS | 20-15 |
| 20-4.1 | BULK MODULUS | 20-15 |
| 20-4.2 | COMPRESSIBILITY | 20-15 |
| 20-4.3 | SPECIFIC GRAVITY | 20-15 |
| 20-4.4 | ELASTIC PROPERTIES OF TUBING | 20-15 |

CHAPTER 1

PROPERTIES OF FEEDBACK CONTROL SYSTEMS***1-1 OBJECTIVES OF A FEEDBACK CONTROL SYSTEM**

The purpose of a feedback control system is to monitor an output (controlled variable) in a manner dictated by an input (reference variable) in the presence of spurious disturbances (such as random load changes). The basic elements of a feedback control system are shown in Fig. 1-1. The system measures the output, compares the measurement with the desired value of the output as prescribed by the input, and uses the error (difference between actual output and desired output) to change the actual output and bring it into closer correspondence with the desired value of the output. To achieve a more sensitive control means, the error is usually amplified; in general, the higher the gain the more accurate the system. Thus, a feedback control system is characterized by measurement, com-

parison, and amplification. In brief, a feedback control system is an error-correcting power-amplifying system that produces a high-accuracy output in accordance with the dictates of a prescribed input.

Since arbitrary disturbances (such as amplifier drift, random torques, etc) can occur at various points in the system, a feedback control system must be able to perform its task with the required accuracy in the presence of these disturbances. Since random noise (unwanted fluctuations) often is present at the input of the system, a feedback control system must be able to reject, or filter out, the noise while producing as faithful a representation of the desired output as is feasible.

**By L. A. Gould*

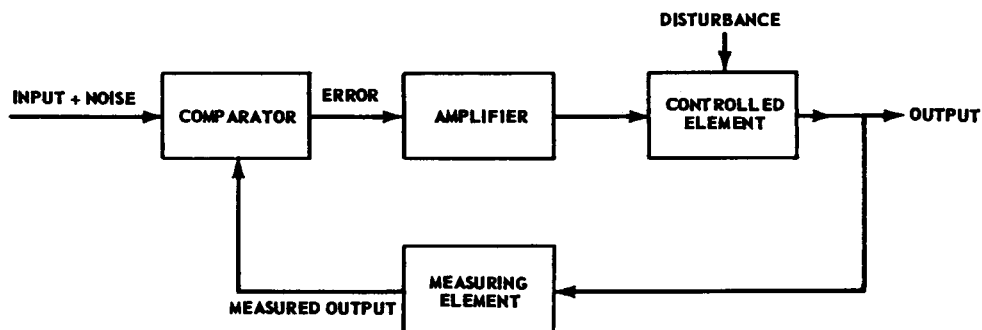


Fig. 1-1 Elements of a feedback control system.

1-2 OPEN-LOOP VS CLOSED-LOOP SYSTEM CHARACTERISTICS

Because a measure of the output is fed back and compared with the input, any representation of a feedback system contains a closed loop (see Fig. 1-1), and the system is thus called a *closed-loop* system. Many control systems do not exhibit this closed-loop feature and may be termed *open-loop* systems. In an open-loop system, the error is reduced by careful calibration. The elements of an open-loop control system are shown in Fig. 1-2.

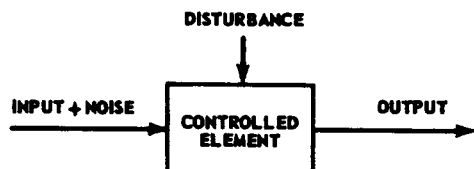


Fig. 1-2 Elements of an open-loop control system.

If open-loop and closed-loop systems are compared, it can be seen that several advantages accrue to the closed-loop system. In a closed-loop system, the percentage change in the response of the system to a given percentage change in the response of one of its elements is approximately inversely proportional to the over-all amplification of the loop. However, in an open-loop system, the percentage change in the response of the system is approximately proportional to the percentage

change in the response of one of its elements. Thus, a feedback control system is insensitive to changes in the parameters of its components and can usually be constructed from less accurate and cheaper components than those used in an open-loop system. One exception to the foregoing statement results from an inherent limitation—the closed-loop system can be no more accurate and reliable than its measuring element. The same limitation holds true for an open-loop system.

The error produced in an open-loop system by a given disturbance is much larger than the error produced by the same disturbance in an equivalent closed-loop system, the ratio of errors being approximately proportional to the over-all amplification of the loop of the closed-loop system. Thus, a feedback control system is relatively insensitive to extraneous disturbances and can be used in situations where severe upsets are expected. One can conclude that the over-all amplification (or gain) that can be achieved inside the loop of a feedback control system directly affects the accuracy of the system, the constancy of its characteristics, and the “stiffness” of the system in the face of external upsets or disturbances. In general, it is found that the higher the gain of the system, the better the system. The highest gain that can be used, however, is limited in every case by considerations of stability.

1-3 STABILITY AND DYNAMIC RESPONSE

For the advantages of accuracy and constancy of characteristics, the feedback control system must pay a price in the form of a greater tendency toward instability. A linear system is said to be stable if the response of the system to any discontinuous input does not exhibit sustained or growing oscillations. Essentially, this means that the system response will ultimately settle down to some steady value. An unstable system which exhibits steady or runaway oscillations is unacceptable. Unstable behavior must be guarded

against in the design, construction, and testing of feedback systems. Because of the possibility of instability, a major portion of control system design is devoted to the task of ensuring that a safe margin of stability exists and can be maintained throughout the operating range of the system.

It can be shown that the cause of instability in a given closed-loop system is due to the fact that no physical device can respond instantaneously to a sudden change at its input. If a sudden change occurs in the error of a

PROPERTIES OF FEEDBACK CONTROL SYSTEMS

feedback system, the output will not correct for the error instantaneously. If the corrective force is great enough (due to a high amplification), the output will accelerate rapidly and cause a reversal of the error. If a high output velocity is attained, the inertia of the output will carry the output past the point where the error is zero. Instability occurs if the maximum magnitude of the error after reversal is equal to, or greater than, the magnitude of the original disturbance in the error. The tendency for a system to become unstable is accentuated as the amplification is increased, since the stored energy in the inertia of the output will be correspondingly increased without any compensating increase in the rate of dissipation of energy in the system. This situation corresponds to an excessive delay in the response of the output. Thus, an attempt to increase accuracy by increasing gain or amplification is usually accompanied by an increased tendency toward instability. As a result, design becomes a compromise between accuracy and stability. A more detailed and quantitative examination of stability

is developed in Ch. 4.

The dynamic behavior of a system is important not only as a determinant of stability but also, for stable systems, as a measure of instantaneous accuracy. In many situations where rapid input variations occur, it is of the utmost importance that the error be kept within specified bounds *at all times*. Ideally, a system with no time lag would be able to follow extremely rapid input variations with perfect accuracy at all times. Actually, the impossibility of achieving instantaneous response, together with the stability problem created by the "pile-up" of the dynamic lags of cascaded elements in a loop, make the problem of maintaining dynamic accuracy (i.e., error within specified bounds *at all times*) progressively more difficult as the rapidity of input variations increases. Consequently, the design of both system and components is focused to a large degree on improving the speed of response (in other words, reducing dynamic lags), thereby obtaining a concomitant improvement in the over-all dynamic accuracy of the system.


1-4 TERMINOLOGY OF FEEDBACK CONTROL SYSTEMS


To facilitate discussion and to maintain uniformity, a specific terminology has been adopted. The general diagram of a feedback control system is shown in Fig. 1-3. Note that some of the elements and variables in this diagram correspond to real devices and signals, whereas other elements and variables correspond to purely hypothetical properties of the system that are useful in visualizing the various functions of the system.

To aid visualization and to distinguish between variables and components, the symbolism of Fig. 1-3 is defined as follows:

(a) A line represents a *variable* or *signal*. The arrow on the line designates the direction of information flow.

(b) A block represents a *device* or *group of devices* that operate on the signal or signals entering the block to produce the signal leaving the block.

(c) The symbol  represents *summation*. The variables entering are added algebraically, according to the signs associated with the corresponding arrows, to produce the variable leaving.

(d) The symbol  is called a *splitting point*. The variable entering is to be transmitted to two points in the diagram. The variables leaving are both *identical* to the variable entering.

The symbolism used for the variables in Fig. 1-3 is defined as follows:

| | |
|-----------------------------------|-------------------------------------|
| r = reference variable or input | u = disturbance or upset |
| n = noise | c = controlled variable or output |
| e = actuating variable | i = desired or ideal output |
| m = manipulated variable | y_e = system error |

THEORY

In many cases, the representation of Fig. 1-3 can be simplified. If the measuring and feedback elements are ideal and have no dynamic lag, it is possible to redraw the figure (see Ch. 3) so as to have no elements in the feedback path of the system.

A system in which the unmodified controlled variable is fed back directly for comparison with the input is called a unity-feedback system. The main loop of a unity-feedback system is shown in Fig. 1-4.

If the ideal output of a system is the refer-

ence variable, the ideal elements are perfect. That is, the desired output is exactly equal to the reference input at all times. Many designs require that the output equal the input at all times although, strictly speaking, this is impossible with real components.

If a unity-feedback system is to have its desired output equal to its reference input, in the absence of noise, the system error must equal the actuating variable. A unity-feedback system of this type is often used initially in the process of design because of its simplicity. Such a system is shown in Fig. 1-5.

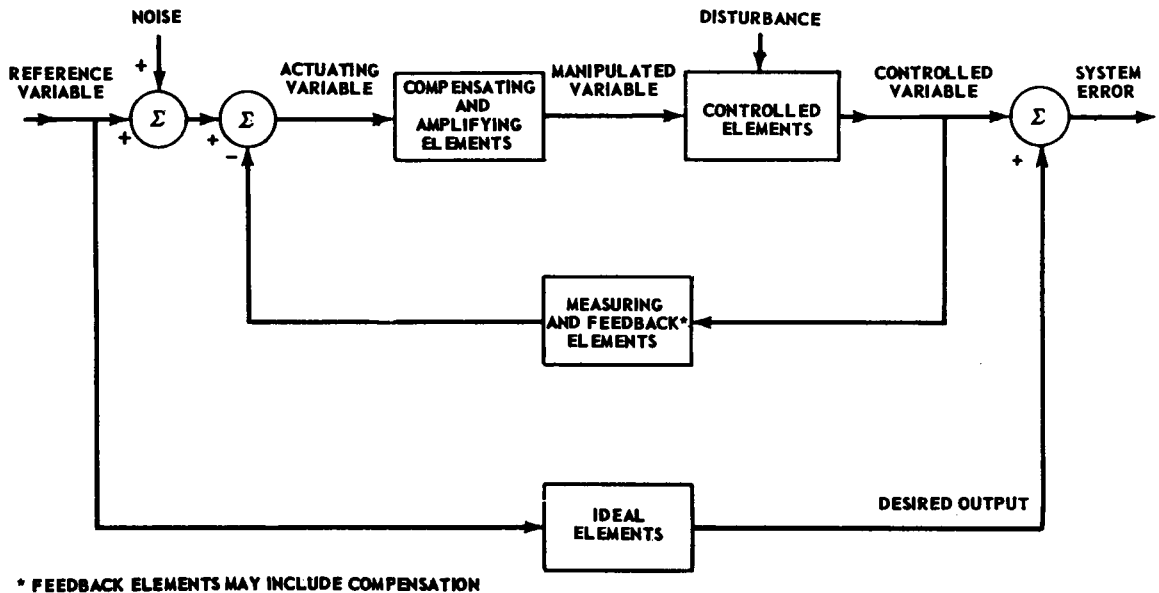


Fig. 1-3 General diagram of a feedback control system.

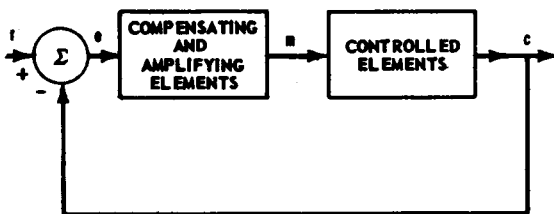


Fig. 1-4 Unity-feedback system main loop.

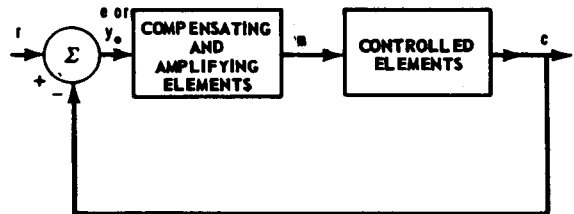


Fig. 1-5 Basic unity-feedback system.

CHAPTER 2

DYNAMIC RESPONSE***2-1 INTRODUCTION**

The *dynamic response* of a component or system is the output response to an input that is a varying function of time. The *steady-state response* of a component or system is the output response to an input that is constant with time.

Paragraph 1-3 indicates that dynamic response is a basic determinant of system stability as well as an important element of system performance. All design theory for a feedback control system is centered on the study, analysis, and manipulation of the dynamic response characteristics of the system and of the components that are part of the system. Because of its fundamental importance, the dynamic response of any physical device or system is classified according to the nature of the input time variation that

occurs. In some cases, the input time variation may be entirely artificial since it may not ordinarily occur in practice (for example, a sinusoidal signal). In other cases, the input variation may be one that is known to occur in practice (for example, a step change). In the former case, the artificial input time function is used primarily to facilitate analysis, design, and testing. In the latter case, the actual response of the system to the known input function is an important measure of performance which both the designer and user need to know in order to verify that the system meets the performance specifications. In either case, a clear understanding of the nature of the input and of the methods for finding the response to it are necessary for successful design.

2-2 LINEARIZATION

The basic tool used to describe the dynamic performance of a device is the set of differential equations that serve as a mathematical model for the actual physical device. Since quantitative techniques are imperative for analysis and design, a mathematical description is necessary. However, when going from the physical device to the differential equation model, one must resort to approximations if usable results are to be expected from a reasonable expenditure of time and effort. If the physical situation is such that it is possible to describe the device with a set of

constant-coefficient linear differential equations to a high degree of accuracy, a wide variety of powerful tools are available to aid analysis. Even when the expected range of variation of the variables is such that the accuracy of approximation is partially lost when constant-coefficient linear differential equations are used, such a representation still serves a useful purpose. Although the representation above is inaccurate, it does provide a qualitative estimate of behavior which is still good enough to be of value to the designer for guiding testing procedures. Furthermore, if the designer artificially restricts the range

*By L. A. Gould

of variation of the variables, he can obtain accurate results which apply to some, though not all, of the expected variations. From such a restriction, there results a partially accurate description that can at least be used to verify whether or not the device meets some of the performance specifications.

Although descriptions utilizing constant-coefficient linear differential equations predominate in feedback control system design, two classes of systems exist that do not lend themselves to such a description. Sampled- or pulsed-data systems are best described by variable-coefficient linear differential equations and are discussed in Ch. 9. Contactor or relay servomechanisms cannot be described by linear equations at all, and one must resort to the nonlinear equations that describe these systems (see Ch. 10). In addition, although the nonlinear properties of linear systems are ordinarily treated as secondary effects in the usual design procedure, they can, under certain circumstances, seriously affect performance. Such circumstances occur when the range of variation of the variables is wide or when the nonlinearity cannot be justifiably ignored. Secondary or incidental nonlinearities such as saturation and backlash are treated in Ch. 10.

Since linearization methods are the basis for most design work, it is important to understand the techniques that are used to establish a linear differential equation description of a device. *As a first step in a linearization procedure, appropriate assumptions are usually formulated based on a knowledge of the physics involved.* For example, in describing d-c machinery operating in an unsaturated region, the effect of hysteresis is often ignored and the normal magnetization curve of the steel is used in the analysis. The justification for such an approximation is based on the fact that, in many cases, the width of the hysteresis loop is small compared with the range of variation of magnetization encountered when using the device. In another situation, one may ignore the effect of backlash in a gear train for reasons analogous to those above.

When reasonable assumptions have eliminated many of the incidental nonlinearities of a device, one is often left with a performance description that is still nonlinear because of the curvature of the steady-state response (steady output as a function of a constant input) curves of the device. When this occurs, use is made of incremental techniques to produce the desired linear description.

The incremental linearization of a nonlinear characteristic (approximate representation of a nonlinear function by a linear function for small changes of the independent variable) is based on the Taylor series expansion of the function around a desired operating point. The deviation of the function from the operating point obtained from the approximate expansion of a function around a steady-state operating point is given by

$$\Delta f(x_1, x_2, \dots, x_n) = C_{x_1} \Delta x_1 + C_{x_2} \Delta x_2 + \dots + C_{x_n} \Delta x_n \quad (2-1)$$

$$\Delta f(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n) - f(x_{10}, x_{20}, \dots, x_{n0}) \quad (2-2)$$

where

$f(x_1, x_2, \dots, x_n)$ = function to be approximated

$(x_{10}, x_{20}, \dots, x_{n0})$ = steady-state operating point

$$\Delta x_k = x_k - x_{k0} \quad (k = 1, 2, \dots, n)$$

$$C_{x_k} = \left. \frac{\partial f}{\partial x_k} \right|_{x_{10}, x_{20}, \dots, x_{n0}} \quad (k = 1, 2, \dots, n)$$

In the approximation [Eq. (2-1)], the deviation or increment of the function from the operating point has been expressed as a linear function of the deviations (from the operating point) of its independent variables. The constant coefficients C_{x_k} are called the partial coefficients of f with respect to x_1, x_2, \dots, x_n at the operating point $(x_{10}, x_{20}, \dots, x_{n0})$.

Example. A shunt d-c motor is speed-controlled by the vacuum-tube circuit shown in Fig. 2-1. Assuming that hysteresis is negligible, the basic equations of the system are as follows:

$$E_p = E, (E_p, I_f) \quad (2-3)$$

DYNAMIC RESPONSE

$$I_f = \frac{E_s - E_p - N_f \frac{dF}{dt}}{R_f} \quad (2-4)$$

$$F = F(I_f) \quad (2-5)$$

$$L_a \frac{dI_a}{dt} + R_a I_a = E_s - E_b \quad (2-6)$$

$$E_b = k_e F N \quad (2-7)$$

$$M = k_m F I_a \quad (2-8)$$

$$M = J \frac{dN}{dt} + M_L \quad (2-9)$$

where

E_p = plate voltage

E_g = grid voltage

I_f = field current

E_s = supply voltage

E_b = motor back emf

L_a = armature inductance

R_a = armature resistance

N = motor shaft speed

k_e = motor back emf constant

N_f = number of field turns

F = field flux

R_f = field resistance

I_a = armature current

M = motor torque

k_m = motor torque constant

J = total moment of inertia

M_L = load torque

It is assumed that mechanical friction is negligible.

To linearize Eq. (2-3), let†

$$E_p \triangleq E_{p0} + e_p = E_{p0} + \Delta E_p \quad (2-10)$$

$$E_g \triangleq E_{g0} + e_g = E_{g0} + \Delta E_g \quad (2-11)$$

$$I_f \triangleq I_{f0} + i_f = I_{f0} + \Delta I_f \quad (2-12)$$

where E_{p0} , E_{g0} , and I_{f0} represent values at the steady-state operating point and e_p , e_g , and i_f represent the deviations of the values of E_p , E_g , and I_f from their corresponding steady-state operating-point values.

†Symbol \triangleq is defined as "equals by definition".

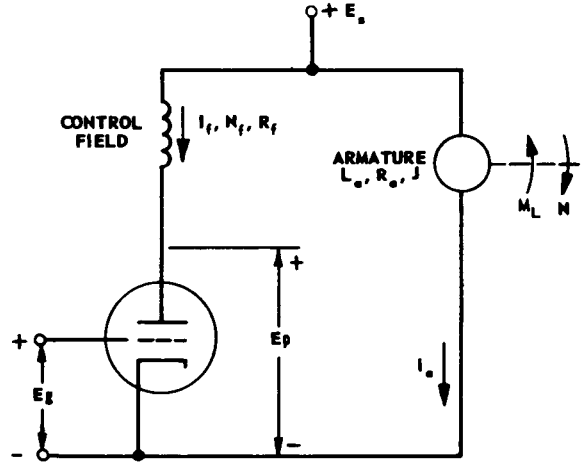


Fig. 2-1 Speed control of shunt d-c motor.

$$\begin{aligned} \text{Since } \Delta E_p &= \left(\frac{\partial E_p}{\partial E_g} \right) \Delta E_g + \left(\frac{\partial E_p}{\partial I_f} \right) \Delta I_f \\ &= \left(\frac{\partial E_p}{\partial E_g} \right) e_g + \left(\frac{\partial E_p}{\partial I_f} \right) i_f, \end{aligned}$$

Eq. (2-10) becomes

$$E_p = E_{p0} + (-\mu) e_g + (R_p) i_f \quad (2-13)$$

where

$$-\mu \triangleq \left. \frac{\partial E_p}{\partial E_g} \right|_{E_{g0}, I_{f0}} \quad \text{and} \quad R_p \triangleq \left. \frac{\partial E_p}{\partial I_f} \right|_{E_{g0}, I_{f0}}$$

As a result of the definitions above, the incremental linear approximation to Eq. (2-3) becomes

$$e_p = -\mu e_g + R_p i_f \quad (2-14)$$

where e_p , e_g , and i_f represent incremental quantities.

To linearize Eq. (2-5), let

$$F \triangleq F_0 + f = F_0 + \Delta F \quad (2-15)$$

Then, Eq. (2-15) can be written as

$$F = F_0 + C_f i_f \quad (2-16)$$

where

$$C_f \triangleq \left. \frac{\partial F}{\partial I_f} \right|_{I_{f0}}$$

THEORY

Consequently, the incremental linear approximation to Eq. (2-5) becomes

$$f = C_f i_f \quad (2-17)$$

Substituting Eqs. (2-10), (2-11), (2-12), (2-15), and (2-17) into Eq. (2-4) results in

$$I_{fo} + i_f = \frac{E_{so} + e_s - (E_{po} + e_p) - N_f \frac{d}{dt}(F_o + C_f i_f)}{R_f} \quad (2-18)$$

or

$$I_{fo} + i_f = \left(\frac{E_{so} - E_{po}}{R_f} \right) + \left(\frac{e_s - e_p - N_f C_f \frac{di_f}{dt}}{R_f} \right) \quad (2-19)$$

where

$$E_s \triangleq E_{so} + e_s$$

$$E_{so} = \text{steady-state value of } E_s$$

$$e_s = \text{increment in } E_s$$

It can be seen that the equation for the operating point of the field circuit is

$$I_{fo} = \frac{E_{so} - E_{po}}{R_f} \quad (2-20)$$

and the incremental equation for the field circuit is

$$i_f = \frac{e_s - e_p - N_f C_f \frac{di_f}{dt}}{R_f} \quad (2-21)$$

To linearize Eq. (2-7), let

$$E_b = E_{bo} + k_e N_o f + k_e F_o n \quad (2-22)$$

where

$$E_b \triangleq E_{bo} + e_b \quad (2-23)$$

$$N \triangleq N_o + n$$

Then, the incremental linear approximation for Eq. (2-7) becomes

$$e_b = k_e (N_o f + F_o n) \quad (2-24)$$

Substituting Eqs. (2-7), (2-17), (2-21), (2-23), and (2-24) into Eq. (2-6), and using a procedure analogous to that above, the equation for the operating point of the armature circuit becomes

$$I_{ao} = \frac{E_{so} - k_e F_o N_o}{R_a} \quad (2-25)$$

and the incremental equation for the armature circuit is

$$L_a \frac{di_a}{dt} + R_a i_a = e_s - k_e N_o C_f i_f - k_e F_o n \quad (2-26)$$

where

$$I_a \triangleq I_{ao} + i_a$$

To linearize Eq. (2-8), use the incremental linear approximation

$$m = k_m F_o i_a + k_m I_{ao} f \quad (2-27)$$

where

$$M \triangleq M_o + m \quad (2-28)$$

By substituting Eqs. (2-23) and (2-28) into Eq. (2-9), the equation for the operating point of the mechanical circuit is

$$M_o = M_{Lo} \quad (2-29)$$

and the incremental equation of the mechanical circuit is

$$m = J \frac{dn}{dt} + m_L \quad (2-30)$$

where

$$M_L \triangleq M_{Lo} + m_L$$

From an examination of the analytical work above, it can be seen that the application of the linearizing technique produces a set of incremental equations that describe the behavior of the system for deviations of the variables from the operating point of the system. In addition, the operating point is also defined by a set of algebraic equations.

DYNAMIC RESPONSE

Summarizing the operating-point equations:

$$I_{fo} = \frac{E_{so} - E_{po}}{R_f} \quad (2-20)$$

$$E_{po} \triangleq E_p(E_{go}, I_{fo}) \quad (2-31)$$

$$I_{ao} = \frac{E_{so} - k_e F_o N_o}{R_a} \quad (2-25)$$

$$F_o \triangleq F(I_{fo}) \quad (2-32)$$

$$M_o = M_{Lo} \quad (2-29)$$

$$M_o \triangleq k_m F_o I_{ao} \quad (2-33)$$

The operating-point equations can be solved for the unknowns I_{fo} , E_{po} , I_{ao} , F_o , N_o , and M_o if the quantities E_{go} , E_{so} , and M_{Lo} are specified. This solution is usually done graphically because the steady-state characteristics of the tube [Eq. (2-31)] and the field structure

[Eq. (2-32)] are presented as experimental curves.

Summarizing the incremental equations:

$$e_p = -\mu e_g + R_p i_f \quad (2-14)$$

$$i_f = \frac{e_s - e_p - N_f C_f \frac{di_f}{dt}}{R_f} \quad (2-21)$$

$$L_a \frac{di_a}{dt} + R_a i_a = e_s - k_e N_o C_f i_f - k_e F_o n \quad (2-26)$$

$$J \frac{dn}{dt} + m_L = k_m F_o i_a + k_m I_{ao} C_f i_f \quad (2-34)$$

The time-varying inputs to the system are the incremental quantities e_g , e_s , and m_L . If these quantities are known, the incremental equations can be solved for e_p , i_f , i_a , and n as functions of time.

2-3 TRANSIENT RESPONSE

The *transient response* of a system is the time variation of one or more of the system outputs following a sudden change in one or more of the system inputs or the derivatives or integrals of the system inputs. Often a transient input variation does not correspond to the actual input variations that a system might experience in practice. However, transient specifications of system performance are very commonly used and, as a result, the designer must know how to describe system behavior in terms of transient response. It can be shown (see Ch. 3) that the transient response of a *linear* system completely specifies the differential equations of the system and, therefore, can be used indirectly to find the response of the system to any type of input.

A given transient response must be referred to the type of input that caused it. Three commonly used transient test inputs are the *impulse*, the *step*, and the *ramp*.

A unit impulse can be conceived of as a time function that is infinite at $t = a$ and zero everywhere else. A unit impulse is defined by Eqs. (2-35) through (2-37), where $\delta_0(t - a)$ is a unit impulse function occurring at $t = a$.

$$\int_{-\infty}^{+\infty} \delta_0(t - a) dt = 1 \quad (2-35)$$

$$\int_{-\infty}^{+\infty} \delta_0(t - a) f(t) dt = f(a) \quad (2-36)$$

$$\delta_0(t - a) = 0, t > a \text{ and } t < a \quad (2-37)$$

The unit step function $\delta_{-1}(t - a)$ is merely the integral of the unit impulse $\delta_0(t - a)$. The unit step is defined by Eqs. (2-38) and (2-39), where $\delta_{-1}(t - a)$ is a unit step occurring at $t = a$.

$$\delta_{-1}(t - a) = \int_{-\infty}^t \delta_0(x - a) dx \quad (2-38)$$

$$\delta_{-1}(t - a) = \begin{cases} 0, & t < a \\ 1, & t > a \end{cases} \quad (2-39)$$

THEORY

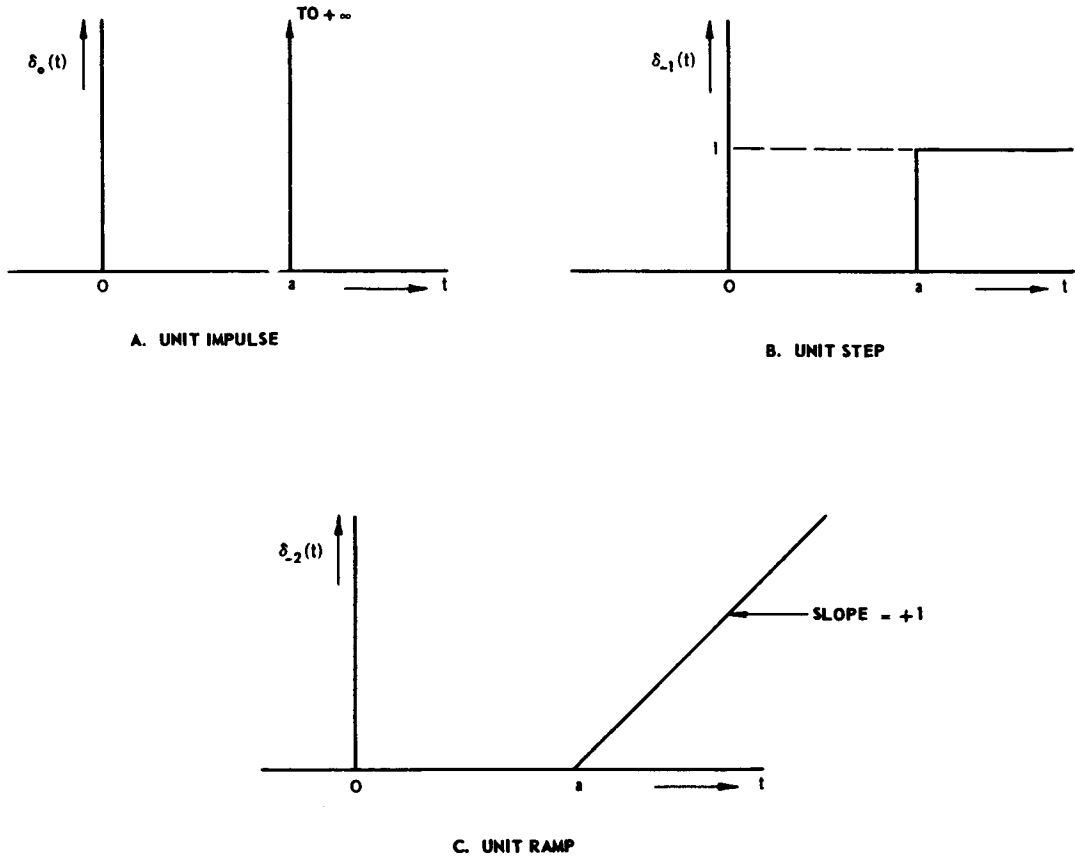


Fig. 2-2 Transient input functions.

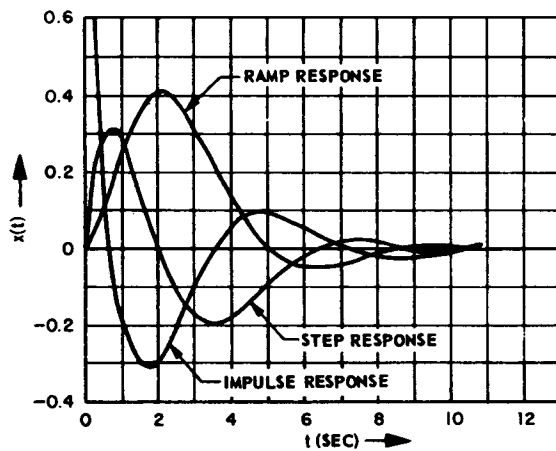


Fig. 2-3 Transient responses for system $\frac{d^3x}{dt^3} + 2 \frac{d^2x}{dt^2} + 2 \frac{dx}{dt} + x = \frac{d^2y}{dt^2}$.

DYNAMIC RESPONSE

The unit ramp function $\delta_{-2}(t - a)$ is the integral of the unit step $\delta_{-1}(t - a)$. The unit ramp is defined by Eqs. (2-40) and (2-41), where $\delta_{-2}(t - a)$ is a unit ramp occurring at $t = a$.

$$\delta_{-2}(t - a) = \int_{-\infty}^t \delta_{-1}(x - a) dx \quad (2-40)$$

$$\delta_{-2}(t - a) = \begin{cases} 0, & t < a \\ t, & t > a \end{cases} \quad (2-41)$$

The impulse, step, and ramp functions are shown in Fig. 2-2. It should be noted that these functions are equal to zero for all $t < a$ and that they are discontinuous or one or more of their derivatives are discontinuous at the instant of occurrence. Clearly, the response of a physical system to any one of these inputs will be zero before the input

occurs if the system is assumed to be initially at rest.

Example. A system is described by the equation

$$\frac{d^3x}{dt^3} + 2 \frac{d^2x}{dt^2} + 2 \frac{dx}{dt} + x = \frac{d^2y}{dt^2} \quad (2-42)$$

where y is the input and x is the output. The output transient responses as functions of time are shown in Fig. 2-3 when the input is a unit impulse, a unit step, and a unit ramp, each occurring at $t = 0$. The specified initial conditions for $t \leq 0$ are $x = 0$, $dx/dt = 0$, and $d^2x/dt^2 = 0$.

It should be noted that, although the curves in Fig. 2-3 are different, they represent the same information about system behavior provided that the input associated with each curve is known.

2-4 FREQUENCY RESPONSE

The *frequency response* of a system is the variation of the output to an input which is a constant-amplitude variable-frequency sinusoid. Frequency response is usually of interest in the linear case but does have application in the nonlinear case (see Ch. 10).

In the case of a linear system, a sinusoidal input produces a sinusoidal output of the same frequency as the input. The frequency response of a linear system is therefore completely described by the ratio of the output amplitude to the input amplitude and by the phase angle of the output relative to the phase angle of the input, both expressed as functions of frequency.

The frequency response of a system is usually presented in three ways: by a polar plot of the tip of the vector $A(\omega) e^{j\phi(\omega)}$ with frequency ω as a parameter ($j = \sqrt{-1}$); by separate plots of $10 \log_{10} A(\omega)$ and $\phi(\omega)$ versus frequency ω ; and by the gain-phase plot of $10 \log_{10} A(\omega)$ versus $\phi(\omega)$ with frequency ω as a parameter. $A(\omega)$ is the amplitude ratio of output to input and $\phi(\omega)$ is the output phase

angle minus the input phase angle. One could also plot $20 \log_{10} A(\omega)$ as is done in the literature in many places, but there is little to be gained by using the factor of two.

The frequency response of a system is useful primarily because of the many theoretical simplifications that are possible when it is used as an analytical and design tool. Just as transient inputs rarely occur in practice, so do sinusoidal inputs almost never occur in practice. Nevertheless, both methods of describing dynamic response are useful in analysis and design.

Since frequency response and transient response are directly related to the differential equation of a system, they contain the same information about system behavior. These two methods of describing dynamic response are merely different approaches to the same end. Both have a useful function to perform in designing control systems. Techniques for correlating the frequency response and transient response of a system are presented in Chs. 3 and 7.

THEORY

2-5 FORCED RESPONSE

The *forced response* of a system is the time response of an output of the system to an arbitrary, but completely defined, variation of one of the system inputs. Forced response is distinguished from transient response in that the input variation associated with the forced response of a system is considered as a continuous time function with no discontinuities in any of its derivatives. A sinusoidal input is a special case of a forcing input which is isolated for special attention because of its theoretical importance.

A typical example of an arbitrary forcing input is the angle of the line-of-sight from a radar antenna to a target that is moving at constant velocity in a straight line (see Fig. 2-4). Such a course is known as a *straight-line crossing course*. The angle of the line-of-sight θ in this case is given by

$$\theta(t) = \tan^{-1} \frac{V}{R} t \quad (2-43)$$

where

V = target velocity

R = minimum target range

The inverse tangent function in Eq. (2-43) and all its derivatives are continuous for all t .

Many design problems have input specifications involving arbitrary forcing functions that cannot be adequately described by discontinuous functions. The techniques for determining the response of a system to these functions are discussed in Pars. 3-3, 3-7, and 7-2.

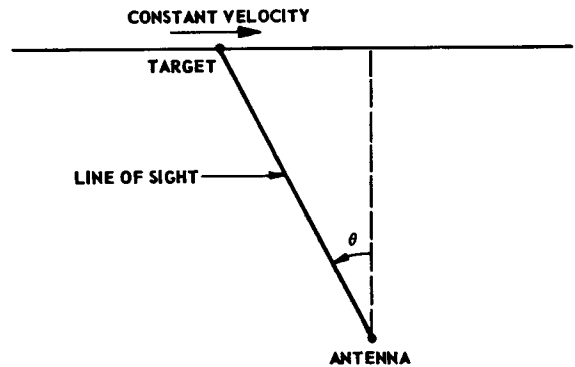


Fig. 2-4 Straight-line crossing course.

2-6 STOCHASTIC INPUTS

A *stochastic process* is one in which there is an element of chance. In many situations, the input to a system is not completely predictable and cannot be described by a mathematical function, either analytically or graphically. The term "random process" is often used to describe such a situation, but it is not an accurate term since a process can often consist of a combination of a completely predictable portion together with a purely random portion. It is evident that the signals in a feedback control system are more often stochastic than predictable in nature, particularly when the effect of the ever-present noise is considered. A typical example of a stochastic process is a radar signal corrupted by noise.

Since a degree of uncertainty exists if one attempts to determine the value of a stochastic signal at a given instant of time, probability density functions and other statistical characterizations such as the average value, the root-mean-square (rms) value, and the correlation function are used to describe the signal. It is useful to think of a stochastic signal as a member of a family of signals, each generated by an identical process. Such a family of signals is called an *ensemble*, and the statistical characterizations of the stochastic process are related to the ensemble of signals rather than to a particular member of the ensemble. Determination of the response of a system to a stochastic input does not yield a function of time, but rather a

DYNAMIC RESPONSE

statistical characterization of the output signal ensemble.

Stochastic signals are separated into two classes. If the statistical behavior of the process that generates the ensemble is independent of time, the process is *stationary*. A *non-stationary* process is one whose statistics vary with time. In most situations involving stochastic processes, the signals generated are non-stationary. It is useful, however, to treat practical processes as stationary if the variation of the statistics with time is small over the useful life of the system. A typical example is the noise generated in a vacuum tube. As the tube ages, the statistical character of the noise changes. If the period of use of the tube is short compared with its expected life, then the noise generated by the tube can be considered as a stationary process.

If a stochastic process is stationary, it is possible to use a single member of the ensemble

of the process to determine the statistics of the process. For example, if the average value of a signal is sought and the process generating the signal is known to be stationary, the average value can be found in two ways. In the first way, the average value is computed by taking the time average for a single member of the ensemble. In the second way, the average value is computed by taking the values of all the members of the ensemble at a particular instant of time and averaging these values. The latter average is called the *ensemble average*. Since the process is known to be stationary, both averages are independent of time. That both averages are identical has not been proven as yet, but their identity seems plausible if one accepts the assumption, the so-called *ergodic hypothesis*, that ensemble averages and time averages are identical for a stationary process. The various ways to characterize stochastic signals and the response of a system to a stochastic signal are discussed in Par. 3-8.

CHAPTER 3

METHODS OF DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS*

3-1 THE DIFFERENTIAL EQUATIONS

As discussed in Ch. 2, any design procedure is based on the differential equations that serve as the mathematical model for the physical system. This chapter deals with methods of determining the dynamic response of physical systems from the differential equations that describe them. The type of response sought depends upon several factors: the specifications of the system; the design procedure adopted; and the limitations imposed by test conditions encountered when seeking experimental verification of the design performance.

Differential equations may be classified as follows:

- (a) Linear differential equations with constant coefficients
- (b) Linear differential equations with time-varying coefficients
- (c) Nonlinear differential equations

Of the three classes, constant-coefficient linear differential equations are, by far, the most widely used and the best understood. The subject matter of this chapter is focused exclusively on methods of solving equations in this class. Chapter 9 deals with time-variable linear differential equations of a specific type that have a wide application. Chapter 10 considers nonlinear equations and some of the techniques for treating them.

The general form of a linear differential equation with constant coefficients is

$$\sum_{i=0}^n a_i \frac{d^i x}{dt^i} = \sum_{j=0}^m b_j \frac{d^j y}{dt^j} \quad (3-1)$$

*By L. A. Gould

where the a 's and b 's are the constant coefficients, $x(t)$ is the response function, and $y(t)$ is the input function. The equation is linear because the response to a sum of component input functions equals the sum of the responses to each of the component input functions. The highest-order derivative of the response x present in the equation is called the *order* of the equation. Thus, Eq. (3-1) is an equation of the n th order. The information necessary for a solution of the equation is a statement of the initial value of the response and the initial values of its first $n-1$ derivatives, as well as the value of the input $y(t)$. The response can be separated into two parts — a general or homogeneous solution, and a particular solution. The complete solution of the differential equation is the sum of the general solution and the particular solution. The general solution always has the form of a sum of exponentials with real and complex arguments; the particular solution has the same form as the input or a sum of the input and its derivatives. The general solution is often called the *force-free* or *transient* solution; the particular solution is called the *forced* or *steady* solution. Each term in the transient solution is called a *normal response mode* or *characteristic* of the equation.

The complete solution of a linear differential equation is given by

$$x(t) = x_p(t) + \sum_{k=1}^n A_k e^{p_k t} \quad (3-2)$$

where $x_p(t)$ is the particular solution, p_k is a root of the equation, and A_k is a constant-amplitude coefficient of a response mode. The

THEORY

root p_k is a function only of the coefficients a_i whereas A_k is a function of the a 's, b 's, and $y(t)$ [Eq. (3-1)]. The quantities A_k and p_k are generally complex numbers that occur in conjugate pairs if the coefficients a_i and b_j and the input function $y(t)$ are real.

The term *root* is applied to each of the p_k 's because these numbers can be found from the differential equation by treating the differentiating operator d/dt as a real variable, replacing it by p for convenience, and setting $y(t)$ equal to zero. The algebraic equation resulting from such substitutions in Eq. (3-1) is

$$\sum_{i=0}^n a_i p^i = 0 \quad (3-3)$$

This equation is known as the *characteristic equation*. The roots of Eq. (3-3), when determined, give the p_k 's of the normal response modes of Eq. (3-2).

The classical procedure for solving constant-coefficient linear differential equations is covered in many textbooks.^(1,2,3,4) More powerful tools for treating differential equations, such as Laplace and Fourier transforms, are presented in Pars. 3-4. For situations where the input is sinusoidal or stochastic, additional special techniques are used. These are discussed later in the text.

3-2 FACTORING AND CHARACTERISTIC PARAMETERS OF RESPONSE MODES

3-2.1 FACTORING

In most cases, the solution of a linear differential equation requires the determination of the roots of the characteristic equation [Eq. (3-3)]. Unfortunately, if the order of the equation is high, the process of factoring the equation to find the roots becomes extremely tedious. For such cases, special techniques have been developed (see Pars. 4-4, 5-7, and 7-1). This section covers some general factoring procedures applicable to any algebraic equation. In addition, the characteristics of first- and second-order equations are discussed and graphical methods for determining the roots of third- and fourth-order equations are presented.

The factoring of rational polynomials is covered by many authors.^(5,6,7,8,9,10,33) The method presented here is one that is very convenient.

The general algebraic equation can be written as

$$f(p) = p^n + c_{n-1}p^{n-1} + \dots + c_1p + c_0 = 0 \quad (3-4)$$

If the order n of the equation is odd, one or more real roots must exist. The real root (or roots) can be determined graphically by plotting $f(p)$ versus p and noting the zero-crossing(s) of $f(p)$, or analytically by using Horner's method of synthetic division, with the first trial divisor being

$$(p + p_1) = p + \frac{c_0}{c_1} \quad (3-5)$$

If the equation is reduced to one of even order, Lin's method⁽⁹⁾ can be used. This involves choosing the trial divisor

$$g_1(p) = p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2} \quad (3-6)$$

Next, $f(p)$ is divided by $g_1(p)$ as follows:

$$\begin{array}{r}
 p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2} \overline{) p^n + c_{n-1}p^{n-1} + \dots + c_1p + c_0} \\
 \underline{\phantom{p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2}} p^n + c_{n-1}p^{n-1} + \dots + c_1p + c_0} \\
 \phantom{p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2}} c'_{n-2}p^{n-2} + c'_{n-3}p^{n-3} + \dots + c'_{n-1}p + c'_0 \\
 \phantom{p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2}} \underline{c'_{n-2}p^{n-2} + c'_{n-3}p^{n-3} + \dots + c'_{n-1}p + c'_0} \\
 \phantom{p^2 + \frac{c_1}{c_2}p + \frac{c_0}{c_2}} \text{remainder}
 \end{array} \quad (3-7)$$

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

If the remainder of Eq. (3-7) is not negligible, a new trial divisor is chosen such that

$$g_2(p) = p^2 + \frac{c'_1}{c'_2} p + \frac{c_0}{c'_2} \quad (3-8)$$

Next, $f(p)$ is divided by the new divisor $g_2(p)$, etc., and the process is continued until

the remainder is negligible. The last divisor is then a factor of the original equation. Then, the quotient [of $f(p)$ and this last divisor] is treated in an identical manner, and the process is repeated until $f(1)$ is factored into quadratic factors whose roots can be determined directly.

Example. Find the roots of the algebraic equation

$$f(p) = p^4 + 10.65p^3 + 89.0p^2 + 15.50p + 27.0 = 0 \quad (3-9)$$

Solution. The first trial divisor is

$$g_1(p) = p^2 + \frac{15.50p}{89.0} + \frac{27.0}{89.0} = p^2 + 0.1742p + 0.303 \quad (3-10)$$

Dividing $f(p)$ by $g_1(p)$ produces

$$\begin{array}{r}
 p^2 + 10.48p + 86.9 \\
 p^2 + 0.1742p + 0.303 \overline{) p^4 + 10.65p^3 + 89.0p^2 + 15.50p + 27.0} \\
 \underline{p^4 + 0.17p^3 + 0.3p^2} \\
 10.48p^3 + 88.7p^2 + 15.50p \\
 \underline{10.48p^3 + 1.8p^2 + 3.18p} \\
 \text{second trial divisor} \longrightarrow 86.9p^2 + 12.32p + 27.0 \\
 \underline{86.9p^2 + 15.14p + 26.3} \\
 \text{remainder} \longrightarrow - 2.82p + 0.7
 \end{array}$$

THEORY

The second trial divisor is

$$g_2(p) = p^2 + \frac{12.32}{86.9}p + \frac{27.0}{86.9} = p^2 + 0.1418p + 0.311 \quad (3-11)$$

Dividing $f(p)$ by $g_2(p)$ yields

$$\begin{array}{r}
 p^2 + 0.1418p + 0.311 \quad \overline{) \begin{array}{l} p^2 + 10.51p + 87.2 \\ p^4 + 10.65p^3 + 89.0p^2 + 15.50p + 27.0 \\ \underline{p^4 + 0.14p^3 + 0.3p^2} \\ 10.51p^3 + 88.7p^2 + 15.50p \\ \underline{10.51p^3 + 1.5p^2 + 3.27p} \\ 87.2p^2 + 12.23p + 27.0 \\ \underline{87.2p^2 + 12.36p + 27.1} \\ -0.13p - 0.1 \end{array} \\
 \text{third trial divisor} \longrightarrow \quad \text{remainder} \longrightarrow
 \end{array}$$

The third trial divisor is

$$g_3(p) = p^2 + \frac{12.23}{87.2}p + \frac{27.0}{87.2} = p^2 + 0.1403p + 0.310 \quad (3-12)$$

Since $g_3(p)$ leaves essentially no remainder, the resulting quadratic factors of $f(p)$ are

$$f(p) = (p^2 + 0.1403p + 0.310)(p^2 + 10.51p + 87.2) \quad (3-13)$$

Factoring the two quadratics in Eq. (3-13), the roots of Eq. (3-9) are found to be

$$p_1, p_2 = -0.0702 \pm j 0.552 \text{ and } p_3, p_4 = -5.26 \pm j 7.72 \quad (3-14)$$

3-2.2 CHARACTERISTIC PARAMETERS OF RESPONSE MODES

3-2.3 First Order: $(p + c_0) = 0$

The response mode corresponding to the first-order factor is of the form

$$x(t) = Ae^{-c_0 t} \quad (3-15)$$

The reciprocal of c_0 is called the *time constant* of the response mode.

Useful plots of exponential functions are presented in Fig. 3-1.

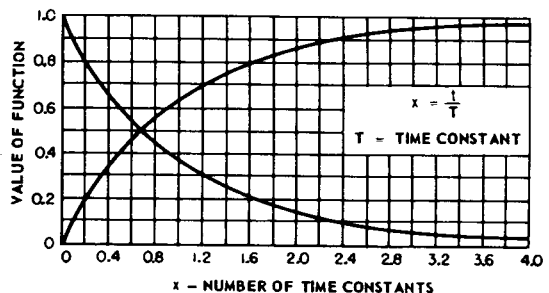


Fig. 3-1 Exponential functions e^{-x} and $1 - e^{-x}$.

3-2.4 Second Order: $p^2 + c_1p + c_0 = 0$

The second-order equation can be rewritten in the form

$$p^2 + 2\zeta\omega_n p + \omega_n^2 = 0 \quad (3-16)$$

where

$$\zeta = \frac{c_1}{2\sqrt{c_0}} = \text{damping ratio}$$

$$\omega_n = \sqrt{c_0} = \text{undamped natural frequency}$$

If $\zeta < 1$, the response mode corresponding to the second-order factor is of the form

$$x(t) = Ae^{-\zeta\omega_n t} \cos(\omega_d t + \phi) \quad (3-17)$$

where

$\omega_d = \omega_n \sqrt{1 - \zeta^2}$ = damped frequency of transient oscillation

If $\zeta \geq 1$, the second-order factor can be factored into two first-order factors so that the response consists of two first-order modes.

3-2.5 Third Order⁽¹¹⁾

$$p^3 + c_2p^2 + c_1p + c_0 = 0$$

By making the substitution

$$p = c_0^{1/3} \lambda$$

the third-order equation is reduced to

$$\lambda^3 + a_2 \lambda^2 + a_1 \lambda + 1 = 0 \quad (3-18)$$

where

$$a_2 = c_2/c_0^{1/3} \text{ and } a_1 = c_1/c_0^{2/3}$$

The reduced equation can be factored as follows:

$$\lambda^3 + a_2 \lambda^2 + a_1 \lambda + 1 = (\lambda + 1/\omega_r^2) (\lambda^2 + 2\zeta\omega_r \lambda + \omega_r^2) \quad (3-19)$$

where

$$a_2 = 2\zeta\omega_r + \frac{1}{\omega_r^2}$$

$$a_1 = \frac{2\zeta}{\omega_r} + \omega_r^2$$

ω_r = a reference frequency

Figure 3-2 shows plots of ω_r versus ζ for constant values of a_2 and a_1 . Figure 3-3 shows plots of a_2 versus a_1 for constant values of ζ and ω_r . From these charts and the third-order equations, the roots of the cubic can be determined.

3-2.6 Fourth Order⁽¹¹⁾:

$$p^4 + c_3p^3 + c_2p^2 + c_1p + c_0 = 0$$

By making the substitution

$$p = c_0^{1/4} \lambda \quad (3-20)$$

the fourth-order equation is reduced to

$$\lambda^4 + a_3 \lambda^3 + a_2 \lambda^2 + a_1 \lambda + 1 = 0 \quad (3-21)$$

where

$$a_3 = c_3/c_0^{1/4}$$

$$a_2 = c_2/c_0^{1/2}$$

$$a_1 = c_1/c_0^{3/4}$$

The reduced equation can be factored into the form

$$(\lambda^2 + 2\zeta_1\omega_{r1}\lambda + \omega_{r1}^2) (\lambda^2 + 2\zeta_2\omega_{r2}\lambda + \omega_{r2}^2) = 0 \quad (3-22)$$

or, alternatively

$$(\lambda^2 + 2\zeta_r\omega_r\lambda + \omega_r^2) [\lambda^2 + 2(\zeta_r\rho_\zeta)(\omega_r\rho_\omega)\lambda + (\omega_r\rho_\omega)^2] = 0 \quad (3-23)$$

where the symbols are defined as follows:

$\omega_r \triangleq \omega_{r1}$ = dimensionless natural frequency of reference component

$\zeta_r \triangleq \zeta_1$ = damping ratio of reference component

$\rho_\omega \triangleq \omega_{r2}/\omega_{r1}$ = ratio of undamped natural frequencies of components

$\rho_\zeta \triangleq \zeta_2/\zeta_1$ = ratio of damping ratios of components

THEORY

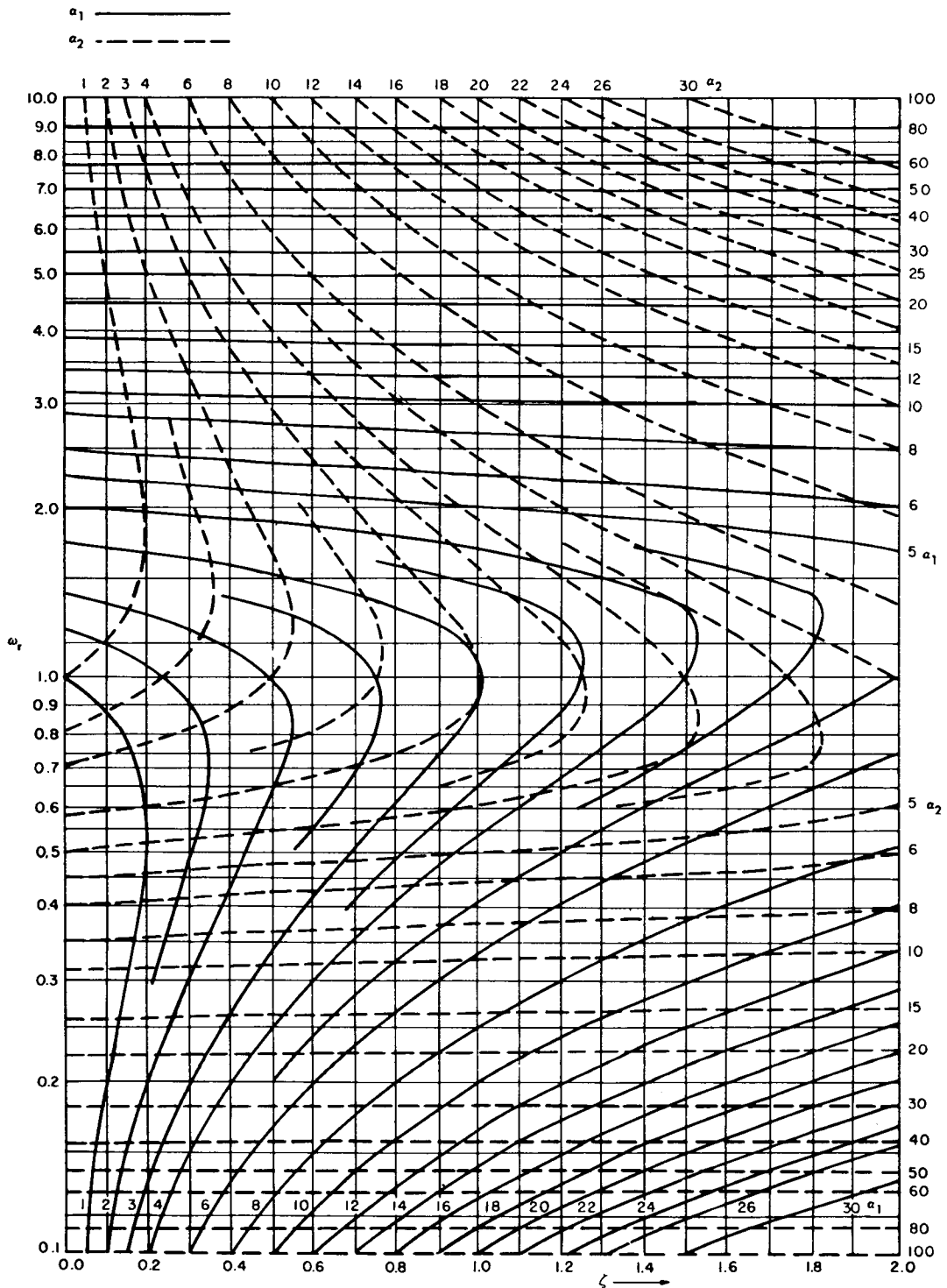


Fig. 3-2 Cubic chart.

By permission from "Solution of the Cubic Equations and the Cubic Charts", by L. W. Evans, bound with "Transient Behavior and Design of Servomechanisms", by G. S. Brown, 1943, Massachusetts Institute of Technology.

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

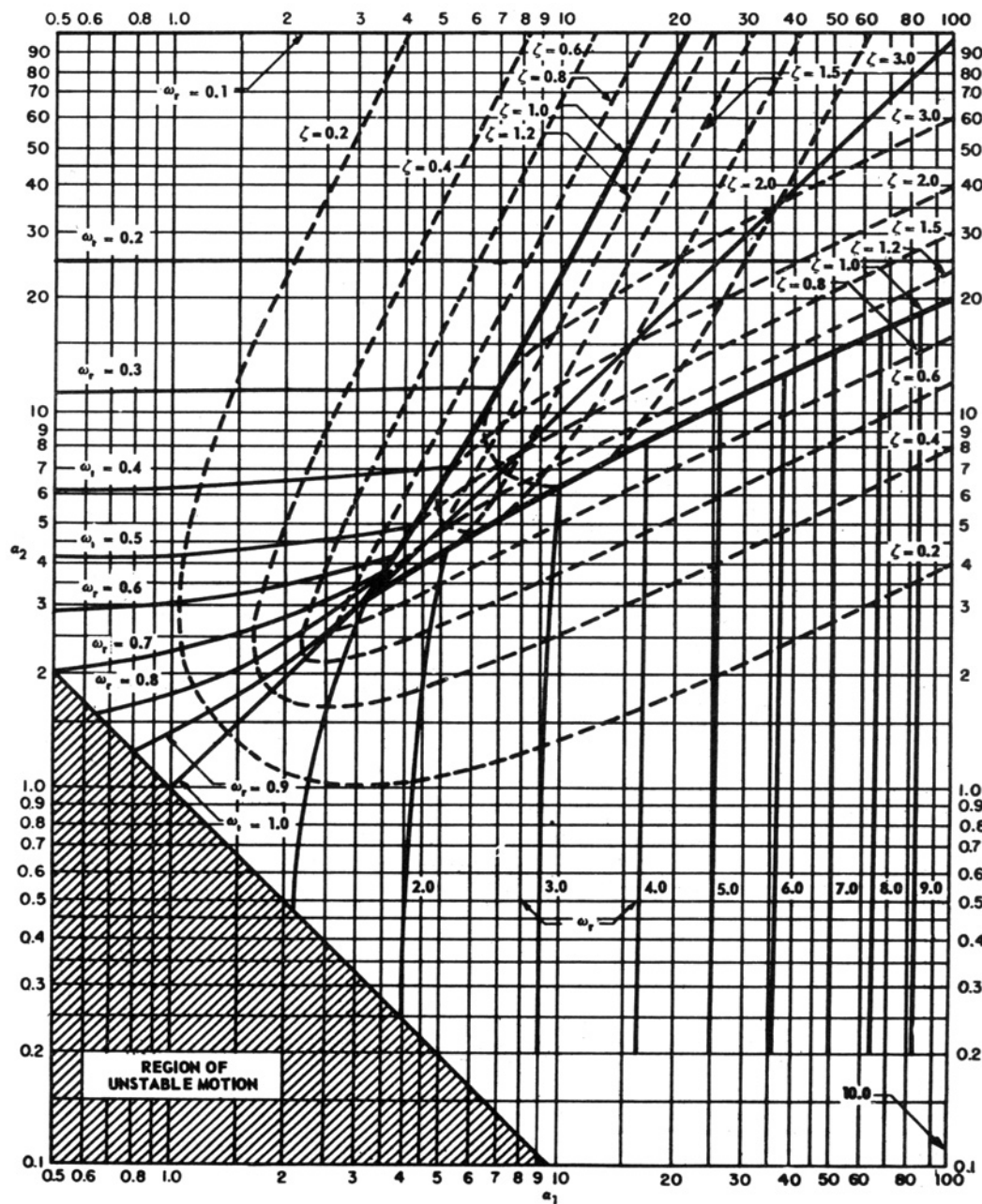


Fig. 3-3 Cubic chart.

By permission from "Solution of the Cubic Equations and the Cubic Charts", by L. W. Evans, bound with "Transient Behavior and Design of Servomechanisms", by G. S. Brown, 1943, Massachusetts Institute of Technology.

THEORY

CHART I FOR EQUATION $\lambda^4 + a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + 1 = 0$

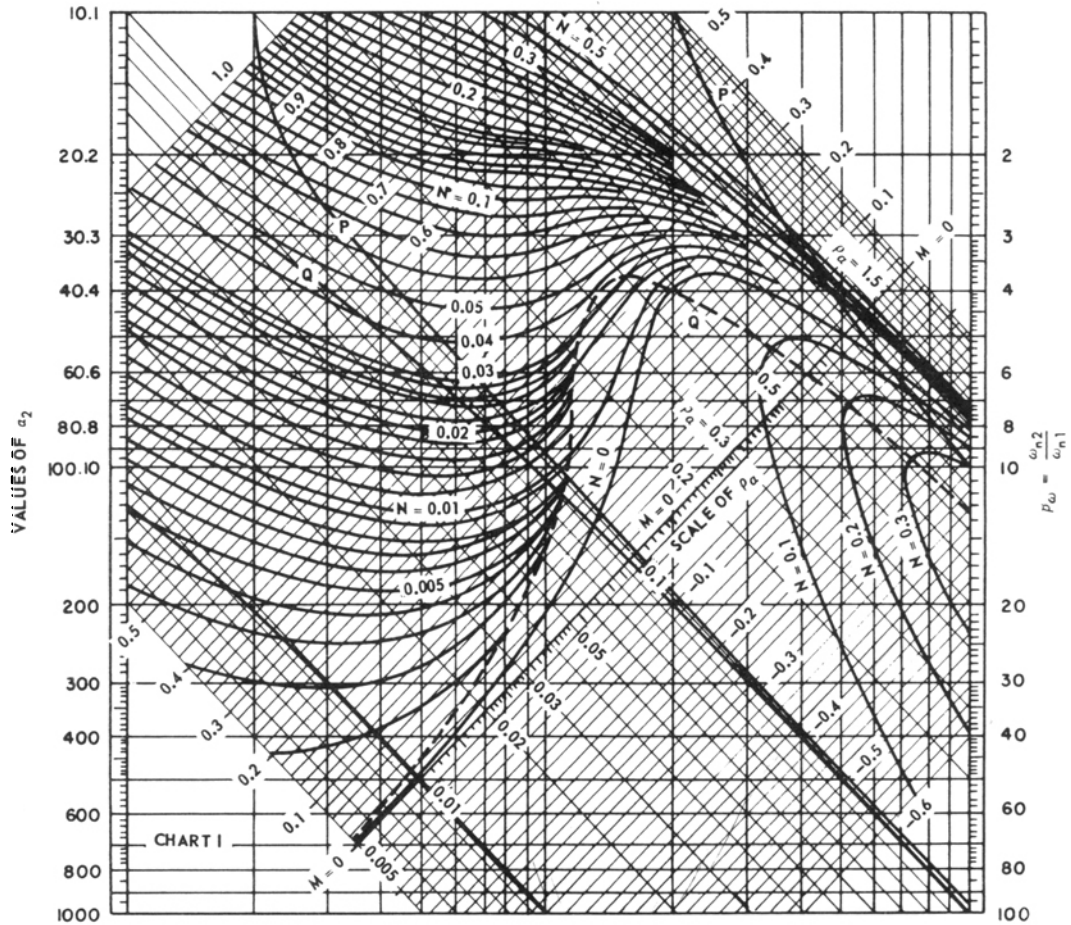


Fig. 3-4 Quartic chart. (Sheet 1 of 2)

By permission from "Servomechanisms", by Y. J. Liu, bound with "Transient Behavior and Design of Servomechanisms", by G. S. Brown, 1943, Massachusetts Institute of Technology.

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

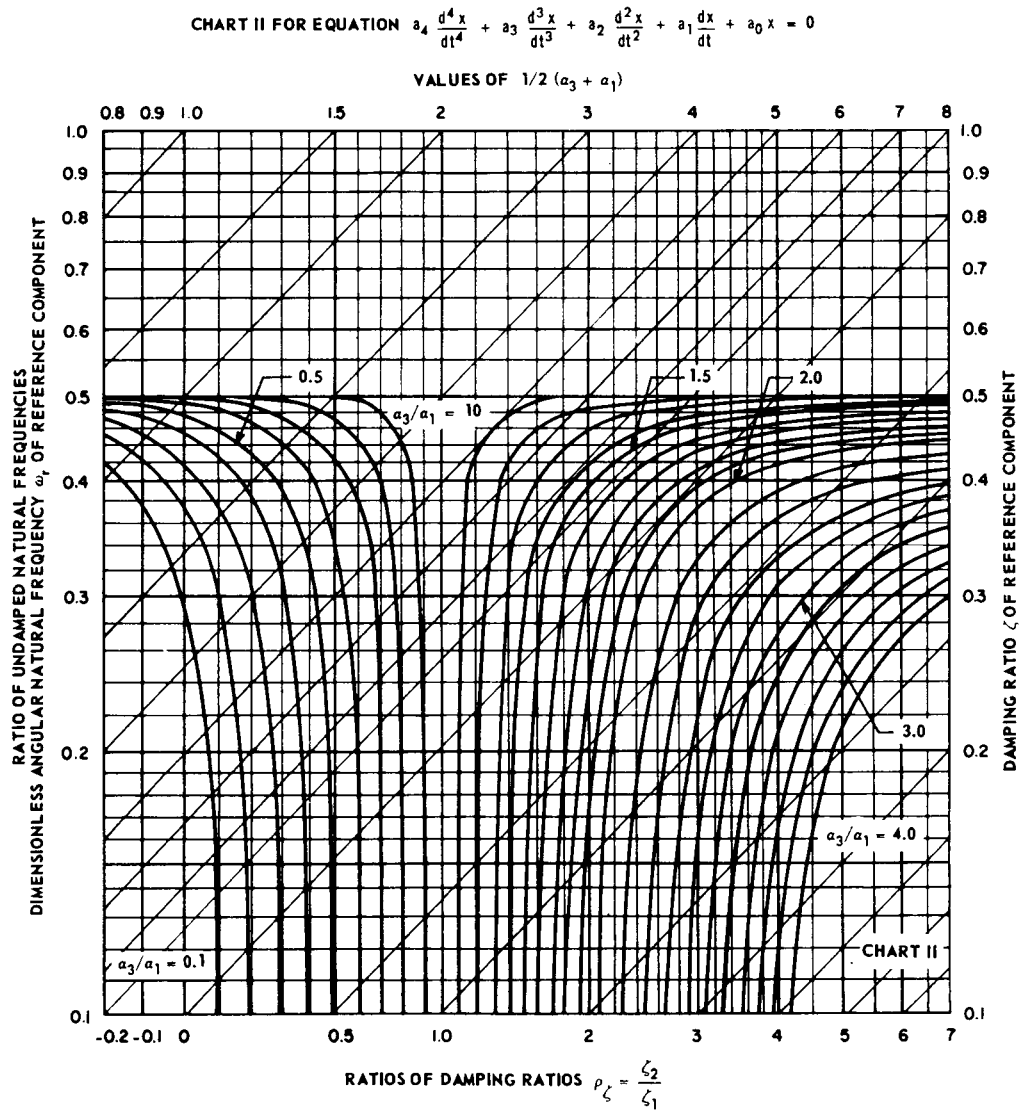


Fig. 3-4 Quartic Chart. (Sheet 2 of 2)

THEORY

To find the parameters defined by Eqs. (3-22) and (3-23), the following procedure is used:

(a) Determine the quantities

$$M \triangleq \frac{\alpha_3 \alpha_{1-4}}{\alpha_2^2}$$

$$N \triangleq \frac{\alpha_3^2 + \alpha_1^2 - 4\alpha_2}{\alpha_2^3}$$

$$\frac{\alpha_3}{\alpha_1} \text{ and } \frac{1}{2} (\alpha_3 + \alpha_1)$$

(b) Stability can be determined from Routh's criterion (Par. 4-2).

(c) The quartic chart is shown in Fig. 3-4.

A sketch of the construction that is used to find ρ_ω , ω_r , ρ_ζ , and ζ_r is shown in Fig. 3-5. Referring to these figures, the determination of the parameters of the factored quartic [Eq. (3-22)] is given by the procedure below.

(d) Locate intersection 3a of the particular pair of M and N values on Chart I. Draw a line through 3a, parallel to the 135°-inclined lines, until it intersects the 45°-inclined scale. The intersection on this scale gives ρ_a , where

$$\rho_a \triangleq \frac{1}{\alpha_2} \left(\rho_\omega + \frac{1}{\rho_\omega} \right)$$

(e) From the particular α_2 value on the left-hand scale, draw a horizontal line until it meets the particular 135°-line found in step (d) at point 3b.

(f) From intersection 3b, draw a vertical line that intersects curve P at 3P and curve Q at 3Q.

(g) A horizontal line drawn through 3P intersects the immediate right scale of ordinates at 3d giving the value of ρ_ω , and the next right (left-hand scale of ordinates of Chart II) at 3d' giving the value of ω_r .

(h) A horizontal line drawn through 3Q on Chart I intersects the particular curve of α_3/α_1 on Chart II at point 4a. The lower abscissa of 4a gives the value of ρ_ζ .

(i) Through 4a, draw a 45°-inclined line until it intersects a vertical line corresponding to the particular value of $(\alpha_3 + \alpha_1)/2$ at point 4b. A horizontal line drawn through 4b intersects the extreme right-hand scale of Chart II giving the value of ζ_r .

(j) When ρ_a is obtained in step (d), the following equations can be used as an alternate method of finding ρ_ω , ω_r , ρ_ζ , and ζ_r :

$$\rho_\omega = \frac{1}{2} \left[\alpha_2 \rho_a + \sqrt{(\alpha_2 \rho_a)^2 - 4} \right] \quad (3-24)$$

$$\omega_r = \frac{1}{\sqrt{\rho_\omega}} \quad (3-25)$$

$$\rho_\zeta = \frac{\rho_\omega \left(\frac{\alpha_3}{\alpha_1} \right) - 1}{\rho_\omega - \left(\frac{\alpha_3}{\alpha_1} \right)} \quad (3-26)$$

$$\zeta_r = \frac{\frac{1}{2} (\alpha_3 + \alpha_1)}{(1 + \rho_\zeta) \left[\sqrt{\rho_\omega} + \frac{1}{\sqrt{\rho_\omega}} \right]} \quad (3-27)$$

$$\zeta_r = \frac{1}{2} \sqrt{\frac{\alpha_2 (1 - \rho_a)}{\rho_\zeta}} \quad (3-28)$$

$$\zeta_r = \frac{\alpha_3}{2 \left[\frac{1}{\sqrt{\rho_\omega}} + \rho_\zeta \sqrt{\rho_\omega} \right]} \quad (3-29)$$

$$\zeta_r = \frac{\alpha_1}{2 \left[\frac{\rho_\zeta}{\sqrt{\rho_\omega}} + \sqrt{\rho_\omega} \right]} \quad (3-30)$$

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

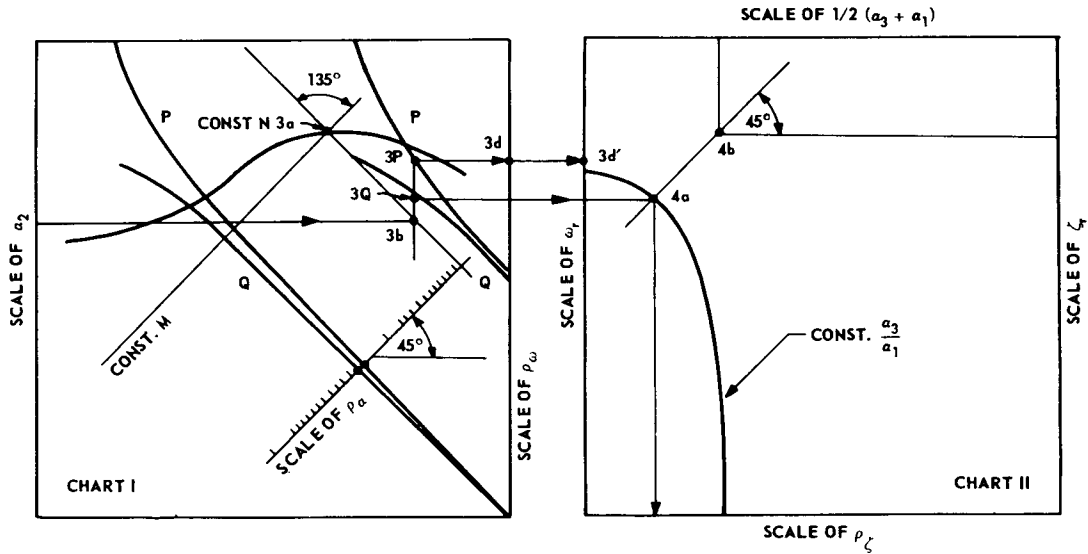


Fig. 3-5 Sketch of the quartic chart.

By permission from "Servomechanisms", by Y. J. Liu, bound with "Transient Behavior and Design of Servomechanisms", by G. S. Brown, 1943, Massachusetts Institute of Technology.

3-3 THE CONVOLUTION INTEGRAL⁽¹²⁾

The output time response of any linear system to an arbitrary input can be found by means of the *convolution* (superposition) integral. If $y(t)$ is the input, $x(t)$ the output, and $w(t)$ the impulse response of the system, then the output x can be found by evaluating the convolution integral

$$x(t) = \int_{-\infty}^{+\infty} dt_1 w(t_1) y(t - t_1) \quad (3-31)$$

or

$$x(t) = \int_{-\infty}^{+\infty} dt_1 w(t - t_1) y(t_1) \quad (3-32)$$

This integral applies in every case and is useful for graphical time-domain studies of system performance. In many situations, however, evaluation of the convolution integral is tedious, so more refined procedures are used (see Pars. 3-4, 3-6, 3-7, 3-8, and 3-9).

If the system being studied is a physical system, then

$$w(t) = 0 \text{ for } t < 0 \quad (3-33)$$

The convolution integral then reduces to

$$x(t) = \int_0^{+\infty} dt_1 w(t_1) y(t - t_1) \quad (3-34)$$

or

$$x(t) = \int_{-\infty}^t dt_1 w(t - t_1) y(t_1) \quad (3-35)$$

If, as often occurs, $y(t)$ and $w(t)$ are both zero for $t < 0$, then the convolution integral reduces to

$$x(t) = \int_0^t dt_1 w(t_1) y(t - t_1) \quad (3-36)$$

or

$$x(t) = \int_0^t dt_1 w(t - t_1) y(t_1) \quad (3-37)$$

THEORY

3-4 LAPLACE AND FOURIER TRANSFORMS^(12,13,14,15,16,17,18,19)

3-4.1 GENERAL

Laplace and Fourier transforms are aids for solving differential equations and introduce properties of system performance that enhance the designer's understanding and simplify his task.

The *Fourier transform* of a function and its inverse are defined as follows:

$$\mathcal{F}[f(t)] \triangleq F(s) \triangleq \int_{-\infty}^{+\infty} dt e^{-st} f(t) \quad [\text{Direct}] \quad (3-38)$$

$$\mathcal{F}^{-1}[F(s)] \triangleq f(t) \triangleq \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} ds e^{st} F(s) \quad [\text{Inverse}] \quad (3-39)$$

where $s = \sigma + j\omega$

The *Laplace transform* of a function and its inverse are defined as follows:

$$\mathcal{L}[f(t)] \triangleq F(s) \triangleq \int_0^{\infty} dt e^{-st} f(t) \quad [\text{Direct}] \quad (3-40)$$

$$\mathcal{L}^{-1}[F(s)] \triangleq f(t) \triangleq \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} ds e^{st} F(s) \quad [\text{Inverse}] \quad (3-41)$$

where s = the complex variable (or frequency) $\sigma + j\omega$.

The Fourier transform is applicable to functions that exist for all time t , whereas the Laplace transform is used for functions that are zero for $t < 0$. In the expression (3-41) for the inverse Laplace transform, the constant c is a convergence factor that enables one to apply the Laplace transform to functions whose Fourier transforms do not exist.

The conditions for the existence of the Fourier transform of a function, known as Dirichlet's conditions, are

(a) $f(t)$ has a finite number of discontinuities in a finite interval

(b) $f(t)$ has a finite number of infinite-valued points in a finite interval

(c) $f(t)$ has a finite number of maxima and minima in a finite interval

(d) $\int_{-\infty}^{+\infty} |f(t)| dt$ is finite

The conditions for the existence of the Laplace transform of a function are identical with those for the Fourier transform except that the fourth condition is relaxed to

$$\int_{-\infty}^{+\infty} |f(t)| e^{-ct} dt \text{ is finite for a finite } c$$

3-4.2 THEOREMS

The following theorems are useful for applying the Laplace and Fourier transforms to the solution of differential equations:

$$(a) \mathcal{L}[af(t)] = aF(s) \quad (3-42)$$

$$(b) \mathcal{L}[f_1(t) \pm f_2(t)] = F_1(s) \pm F_2(s) \quad (3-43)$$

$$(c) \mathcal{L}\left[\frac{d^n f(t)}{dt^n}\right] = s^n F(s) - s^{n-1}f(0+) - s^{n-2}f'(0+) - \dots - sf^{(n-2)}(0+) - f^{(n-1)}(0+) \quad (3-44)$$

$$(d) \mathcal{L}\left[\int_{-\infty}^t \dots \int_{-\infty}^t f(t) dt\right] = \frac{F(s)}{s^n} + \frac{\int_{-\infty}^{0+} f(t) dt}{s^n} + \frac{\int_{-\infty}^{0+} \int_{-\infty}^t f(t) dt}{s^{n-1}} + \dots + \frac{\int_{-\infty}^{0+} \left[\int_{-\infty}^t \dots \int_{-\infty}^t f(t) dt\right]}{s} \quad (3-45)$$

$$(e) \mathcal{L}\left[f\left(\frac{t}{a}\right)\right] = aF(as) \quad (3-46)$$

$$(f) \mathcal{L}\left[\int_0^t f_1(t-\tau)f_2(\tau) d\tau\right] = F_1(s)F_2(s) \quad (3-47)$$

$$(g) \mathcal{L}[f(t-a)] = e^{-as}F(s) \text{ if } f(t-a) = 0 \text{ for } 0 < t < a \quad (3-48)$$

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

$$\begin{aligned} \text{(h)} \quad \mathcal{L}[f(t+a)] &= e^{as}F(s) \\ \text{if } f(t+a) &= 0 \text{ for } -a < t < 0 \end{aligned} \quad (3-49)$$

$$\text{(i)} \quad \lim_{s \rightarrow 0} s\mathcal{L}[f(t)] = \lim_{t \rightarrow \infty} f(t) \quad (3-50)$$

$$\text{(j)} \quad \lim_{s \rightarrow \infty} s\mathcal{L}[f(t)] = \lim_{t \rightarrow 0} f(t) \quad (3-51)$$

$$\begin{aligned} \text{(k)} \quad \mathcal{L}[f_1(t)f_2(t)] \\ = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_1(s-w)F_2(w) dw \end{aligned} \quad (3-52)$$

$$\text{(l)} \quad \mathcal{L}[f_1(t)f_2(t)] \neq F_1(s)F_2(s) \quad (3-53)$$

Theorems (a), (b), (e), (f), (i), (k), and (l) also apply to the Fourier transform. Theorem (c) is called the *real differentiation* theorem and theorem (d) is called the *real integration* theorem. Theorem (e) is used to change the time scale of a problem and is called the *normalization* theorem. Theorem (f) is the *real convolution* theorem and, if applied to the convolution integral [Eq. (3-35)], yields the very important result

$$\begin{aligned} \mathcal{F}[x(t)] &= X(s) = \mathcal{F}\left[\int_{-\infty}^t dt_1 w(t-t_1) y(t_1)\right] \\ &= W(s)Y(s) \end{aligned} \quad (3-54)$$

Theorems (g) and (h) apply to the Fourier transform without the stated restrictions. Theorem (i) is called the *final-value* theorem and theorem (j) is called the *initial-value* theorem. Theorems (k) and (l) are included primarily to prevent the common error suggested by theorem (l), namely, *incorrectly* stating that the transform of the product of two time functions is the product of the separate transforms of the functions.

3-4.3 SOLUTION OF DIFFERENTIAL EQUATIONS

The solution of ordinary linear differential equations is accomplished by means of theorems (a), (b), (c), and (d). Applying these theorems to Eq. (3-1), one obtains

$$\begin{aligned} \left[\sum_{i=0}^n a_i s^i \right] X(s) - A(s) &= \left[\sum_{j=0}^m b_j s^j \right] \\ &Y(s) + B(s) \end{aligned} \quad (3-55)$$

where $A(s)$ is a polynomial in s depending upon the a 's and the initial values of x and its first $(n-1)$ derivatives, and $B(s)$ is a polynomial in s depending upon the b 's and the initial values of y and its first $(m-1)$ derivatives. The response transform can be obtained by solving Eq. (3-55) for $X(s)$

$$X(s) = \left[\frac{\sum_{j=0}^m b_j s^j}{\sum_{i=0}^n a_i s^i} \right] Y(s) + \left[\frac{B(s) + A(s)}{\sum_{i=0}^n a_i s^i} \right] \quad (3-56)$$

In words, this equation can be written

$$\begin{aligned} \left(\begin{array}{c} \text{response} \\ \text{transform} \end{array} \right) &= \left(\begin{array}{c} \text{system} \\ \text{function} \end{array} \right) \left(\begin{array}{c} \text{input} \\ \text{transform} \end{array} \right) \\ &+ \left(\begin{array}{c} \text{initial condition} \\ \text{function} \end{array} \right) \end{aligned} \quad (3-57)$$

The ratio of the response transform to the input transform when all initial conditions are zero (i.e., the initial condition function is zero) is called the *system function* or the *transfer function* of the system. This function depends only upon the coefficients of the differential equation and is independent of the input and the initial conditions. Comparing Eq. (3-57) (with initial condition function set equal to zero) with Eq. (3-54), it is evident that the transfer function of a system equals the transform of the impulse response of the system for a unit impulse.

Transforming a differential equation enables the analyst to replace the processes of differentiation and integration by simple algebraic processes. Then, the transform $X(s)$ can be found algebraically. Subsequently, the system response $x(t)$ corresponding to the response transform $X(s)$ can be found by using the inversion theorem [Eq. (3-41)]. However, this theorem usually involves contour integration in the complex s plane. To avoid this integration, tables of transform pairs have been constructed that give the time function corresponding to a given transform directly. A brief list of commonly used transform pairs is given in Table 3-1. More extensive tables can be found in references (13), (20), and (21).

THEORY

TABLE 3-1 LAPLACE TRANSFORM PAIRS

| $F(s)$ | $f(t), t \geq 0$ |
|---|--|
| 1 | $\delta_0(t)$, unit impulse |
| $\frac{1}{s}$ | $\delta_{-1}(t)$, unit step |
| $\frac{1}{s^2}$ | $\delta_{-2}(t)$, unit ramp |
| $\frac{1}{Ts + 1}$ | $\frac{1}{T} e^{-t/T}$ |
| $\frac{\omega}{s^2 + \omega^2}$ | $\sin \omega t$ |
| $\frac{s}{s^2 + \omega^2}$ | $\cos \omega t$ |
| $\frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2}$ | <p>(1) $\zeta < 1$: $\frac{1}{\omega_n \sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1 - \zeta^2} t$</p> <p>(2) $\zeta = 1$: $t e^{-\omega_n t}$</p> <p>(3) $\zeta > 1$: $\frac{1}{\omega_n \sqrt{\zeta^2 - 1}} e^{-\zeta\omega_n t} \sinh t \omega_n \sqrt{\zeta^2 - 1}$</p> |
| $\frac{1}{(s + \alpha)^2 + \beta^2}$ | $\frac{1}{\beta} e^{-\alpha t} \sin \beta t$ |
| $\frac{s + \alpha}{(s + \alpha)^2 + \beta^2}$ | $e^{-\alpha t} \cos \beta t$ |
| $\frac{1}{s^n}$ | $\frac{1}{(n - 1)!} t^{n-1}$ |
| $\frac{1}{(Ts + 1)^n}$ | $\frac{1}{(n - 1)!} \frac{t^{n-1}}{T^n} e^{-t/T}$ |

If tables of transform pairs are unavailable, or if the particular transform whose inverse is sought is not listed in the tables, the method of partial fractions may be used to expand the transform into a sum of terms, each of which is readily recognized as the transform of a simple time function. If the transform whose inverse is sought is a ratio of rational polynomials, the roots of the numerator polyno-

mial are called the *zeros* of the function and the roots of the denominator polynomial are called the *poles* of the function. If the poles of the function are not repeated, they are called *single-order* poles. The order of a *multiple-order* pole is the number of times the pole is repeated. For a function containing only single-order poles, the partial-fraction expansion of the function is

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

$$F(s) \triangleq \frac{N(s)}{D(s)} = \sum_{k=1}^n \frac{K_k}{s - s_k} \quad (3-58)$$

where

$$K_k \triangleq \left[\frac{(s - s_k) N(s)}{D(s)} \right]_{s=s_k} \\ = \left[\frac{N(s)}{D'(s)} \right]_{s=s_k} \quad (3-59)$$

and s^k is the k th root of the denominator polynomial $D(s)$.

If the transform contains multiple-order poles, the partial-fraction expansion of the function is

$$F(s) \triangleq \frac{N(s)}{D(s)} = \sum_{k=1}^n \sum_{j=1}^{m_k} \frac{K_{kj}}{(s - s_k)^{m_k-j+1}} \quad (3-60)$$

where

$$K_{kj} \triangleq \frac{1}{(j-1)!} \\ \left\{ \frac{d^{j-1}}{ds^{j-1}} \left[\frac{(s - s_k)^{m_k} N(s)}{D(s)} \right] \right\}_{s=s_k} \quad (3-61)$$

and m_k is the order of the pole of $F(s)$ at $s = s_k$.

From Eqs. (3-58) and (3-60), it is obvious that the expansion of a rational function when inverted produces a sum of exponential terms for the corresponding time function. Terms containing exponentials with complex arguments will appear in conjugate pairs and can therefore be combined to form product terms (exponential multiplied by a sine or cosine function) representing damped sinusoids.

Example. The system defined by the equation

$$\frac{d^4 x}{dt^4} + 10.65 \frac{d^3 x}{dt^3} + 89.0 \frac{d^2 x}{dt^2} + 15.50 \frac{dx}{dt} \\ + 27.0x = 27.0y \quad (3-62)$$

is initially at rest. At $t = 0$, a unit ramp input is applied. Find the difference between the input y and the output x as a function of time.

Solution. Since the system is initially at rest, all initial conditions are zero. Transforming Eq. (3-62) results in

$$X(s) = \\ \frac{27.0}{s^4 + 10.65s^3 + 89.0s^2 + 15.50s + 27.0} Y(s) \quad (3-63)$$

Let

$$e(t) = y(t) - x(t) \quad (3-64)$$

Then, transforming Eq. (3-64) and substituting for $X(s)$ from Eq. (3-63), $E(s)$ becomes

$$E(s) = \\ \frac{s[s^3 + 10.65s^2 + 89.0s + 15.50]}{s^4 + 10.65s^3 + 89.0s^2 + 15.50s + 27.0} Y(s) \quad (3-65)$$

By referring to the values of the roots given in Eq. (3-14), the denominator $D(s)$ of Eq. (3-65) can be factored as follows:

$$D(s) = (s + 0.0702 - j0.552)(s + 0.0702 + j0.552) \\ (s + 5.26 - j7.72)(s + 5.26 + j7.72) \quad (3-66)$$

THEORY

The transform of $y(t)$, found from Table 3-1, is

$$Y(s) = \frac{1}{s^2} \quad (3-67)$$

Using the factored form of the denominator and substituting $1/s^2$ for the value of $Y(s)$ in Eq. (3-65) results in

$$E(s) = \frac{[s^3 + 10.65s^2 + 89.0s + 15.50]}{s[s + 0.0702 - j0.552][s + 0.0702 + j0.552][s + 5.26 - j7.72][s + 5.26 + j7.72]} \quad (3-68)$$

Since two pairs of the poles of $E(s)$ appear as conjugate pairs, the partial-fraction expansion of $E(s)$ can be written

$$E(s) = \frac{K_1}{s} + \frac{K_2}{s + 0.0702 - j0.552} + \frac{\overline{K_2}}{s + 0.0702 + j0.552} + \frac{K_3}{s + 5.26 - j7.72} + \frac{\overline{K_3}}{s + 5.26 + j7.72} \quad (3-69)$$

where a bar over a constant indicates the complex conjugate of the constant. Using the expansion theorem [Eq. (3-59)]

$$K_1 = \frac{15.50}{27.0} = 0.574 \quad (3-70)$$

$$K_2 = \left[\frac{(s^3 + 10.65s^2 + 89.0s + 15.50)}{s(s + 0.0702 + j0.552)(s^2 + 10.51s + 87.2)} \right]_{s = -0.0702 + j0.552} = 0.918 e^{-j0.602\pi} \quad (3-71)$$

$$\overline{K_2} = 0.918 e^{+j0.602\pi} \quad (3-72)$$

$$K_3 = \left[\frac{(s^3 + 10.65s^2 + 89.0s + 15.50)}{s(s^2 + 0.1403s + 0.310)(s + 5.26 + j7.72)} \right]_{s = -5.26 + j7.72} = 2.89 \times 10^{-4} e^{-j0.277\pi} \quad (3-73)$$

$$\overline{K_3} = 2.89 \times 10^{-4} e^{+j0.277\pi} \quad (3-74)$$

Inverse transforming Eq. (3-69)

$$\begin{aligned}
 e(t) &= 0.574 + 0.918 e^{-0.0702t} \\
 &\quad [e^{j(0.552t - 1.89)} + e^{-j(0.552t - 1.89)}] \\
 &\quad + 2.89 \times 10^{-4} e^{-5.26t} \\
 &\quad [e^{j(7.72t - 0.87)} + e^{-j(7.72t - 0.87)}] \quad (3-75)
 \end{aligned}$$

The bracketed functions on the right side of Eq. (3-75) are recognized as cosine functions, so that $e(t)$ can be written as

$$\begin{aligned}
 e(t) &= 0.574 + 1.836 e^{-0.0702t} \\
 &\quad \cos(0.552t - 1.89) \\
 &\quad + 5.78 \times 10^{-4} e^{-5.26t} \cos(7.72t - 0.87) \quad (3-76)
 \end{aligned}$$

It is convenient for plotting purposes to write the arguments of the cosine functions in degrees and to use trigonometric identities to reduce the phase angles to angles smaller than 45° . If this is done, $e(t)$ can be written as

$$\begin{aligned}
 e(t) &= 0.574 + 1.836 e^{-0.0702t} \sin \\
 &\quad (31.6t - 18.3)^\circ - 5.78 \times 10^{-4} e^{-5.26t} \sin \\
 &\quad (442t + 40.2)^\circ \quad (3-77)
 \end{aligned}$$

3-4.4 FREQUENCY RESPONSE

It is often important to find the response of a system to a sinusoidal input. For a sinusoidal input, the output of the system will also be sinusoidal after transients have died out. The amplitude and phase angle of the output relative to the input are dependent only upon $W(s)$, the transfer function of the system, and can be determined by letting $s = j\omega$ in the transfer function, where ω is the frequency (in radians/second) of the input sinusoid. The ratio of output amplitude to input amplitude is then given by

$$\frac{A_x}{A_y} = |W(j\omega)| \quad (3-78)$$

where A_x is the output amplitude, A_y is the input amplitude, and $W(s)$ is the transfer function of the system. The phase angle of the output relative to the phase angle of the input is given by†

$$\phi_x - \phi_y = \angle W(j\omega) \quad (3-79)$$

where ϕ_x is the output phase angle and ϕ_y is the input phase angle.

When the transfer function of a system is evaluated as a function of frequency for a sinusoidal input, the complex function that results is called the *frequency response* of the system.

†Symbol \angle denotes "angle"

3-5 BLOCK DIAGRAMS AND SIGNAL-FLOW GRAPHS (22,23,24,25,26,27,28)

3-5.1 BLOCK DIAGRAMS

Equations (3-54) and (3-57) demonstrate that, with zero initial conditions, the transform of the output of a system can be expressed in terms of the input transform and the system function. The system function can be thought of as an *operator*. That is, the system function operates on the input transform to produce the output transform. In a similar manner, the system operates on the input to

produce the output in the time domain, the operation being defined by the convolution integral [Eq. (3-31)] and depending only upon the impulse response of the system. The concept of an operator is presented pictorially by the technique known as *operational block diagram algebra*. The block diagram of a system is the pictorial representation of the mathematical operations involved in the differential equations that describe the system.

THEORY

Table 3-2 presents a list of symbols used in the block diagram representation of a system and Fig. 3-6 presents a list of reductions that enable one to simplify or reduce the block diagrams of a system. Since the block diagram contains no more information than the differential equations, the manipulation of a block diagram is merely a pictorial process of manipulating the differential equations. The advantage of a block diagram representation is that the operational relations in a system are emphasized rather than the hardware. By becoming familiar with common block arrangements, the designer can interpret the function of various elements in a system much more rapidly than would be possible from an inspection of the differential equations.

Example. The transformed equations of a servomotor driving an inertia load coupled to the motor through a flexible shaft are

$$T_m = (J_m s^2 + f_m s) \theta_m + K(\theta_m - \theta_L) \quad (3-80)$$

$$K(\theta_m - \theta_L) = J_L s^2 \theta_L + T_L \quad (3-81)$$

where

T_m = motor torque K = shaft stiffness

J_m = motor inertia θ_L = load angle


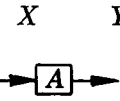
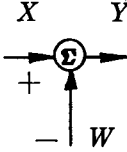
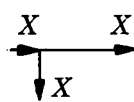
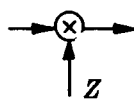
f_m = motor damping J_L = load inertia

θ_m = motor angle T_L = load torque

The damping of the flexible shaft is assumed to be negligible. Draw the block diagram of the system and reduce the diagram, keeping the motor angle θ_m and the load angle θ_L in evidence.

Solution. The block diagram of the system is drawn in its "primitive" form in Fig. 3-7A. The successive steps necessary to reduce the "primitive" diagram to the desired form are shown in Figs. 3-7B to 3-7I with the rules used for each step indicated below each step.

TABLE 3-2 BLOCK DIAGRAM SYMBOLS

| Symbol | Description | Operation |
|---|-----------------|-------------|
|  | variable | |
|  | operator | $Y = AX$ |
|  | summing point | $Y = X - W$ |
|  | splitting point | $X = X$ |
|  | multiplier | $Y = XZ$ |

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

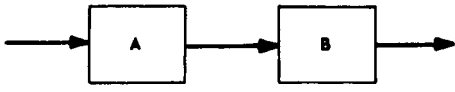
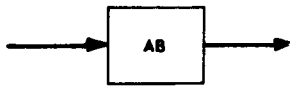
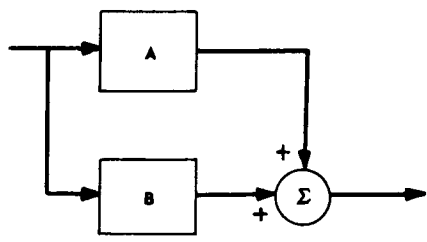
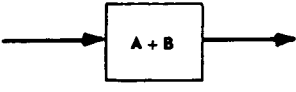
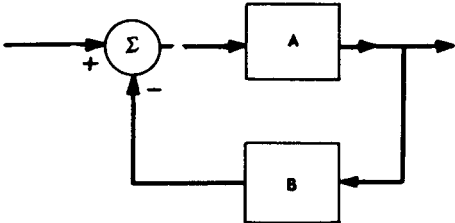
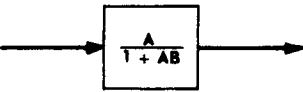
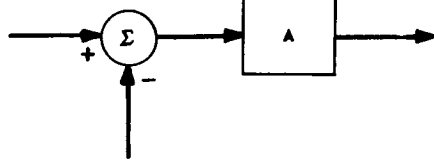
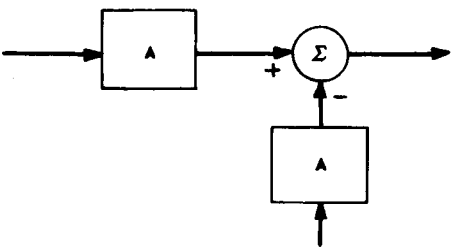
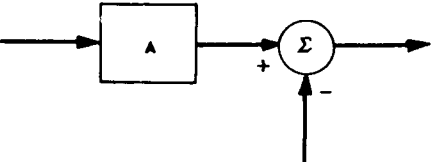
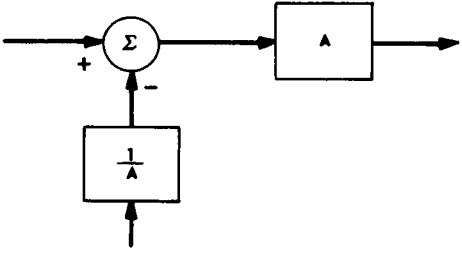
| RULE | ORIGINAL DIAGRAM | EQUIVALENT DIAGRAM |
|------|---|--|
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

Fig. 3-6 Block diagram manipulation and reduction "rules". (Sheet 1 of 3)

THEORY

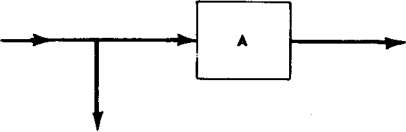
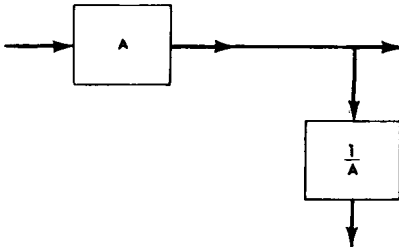
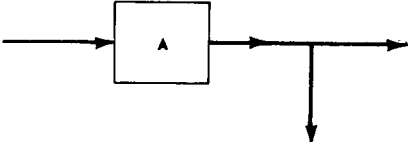
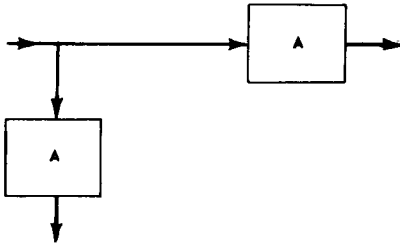
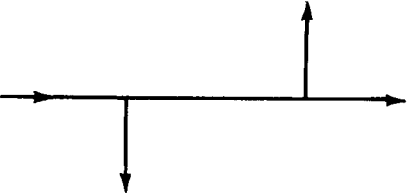
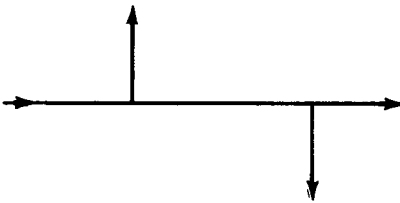
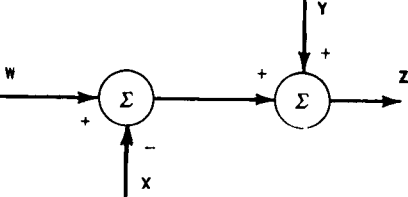
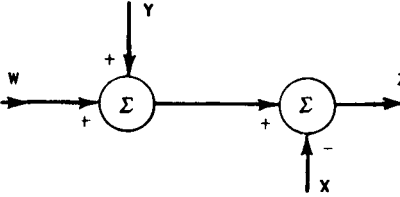
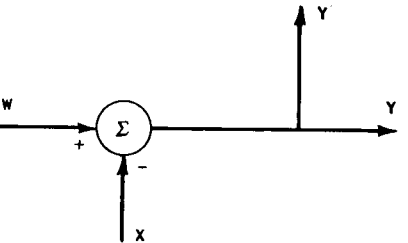
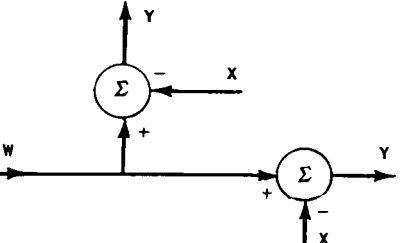
| RULE | ORIGINAL DIAGRAM | EQUIVALENT DIAGRAM |
|------|---|--|
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |

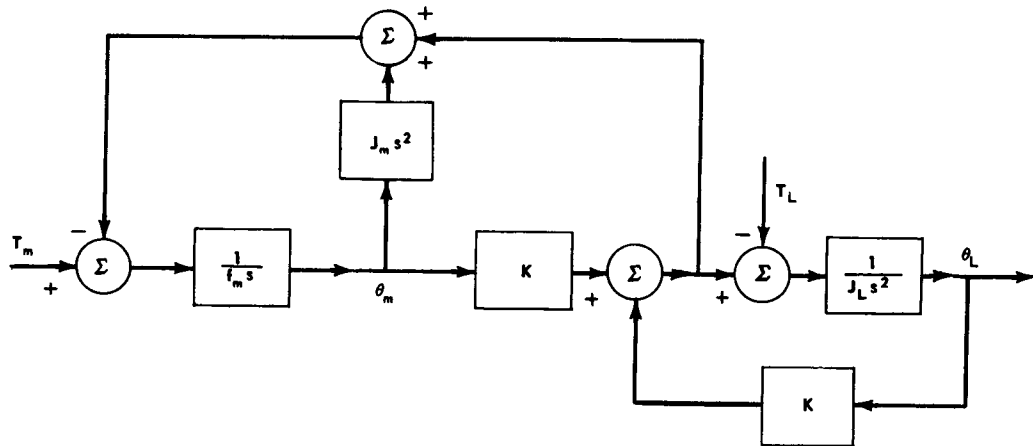
Fig. 3-6 Block diagram manipulation and reduction "rules". (Sheet 2 of 3)

DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS

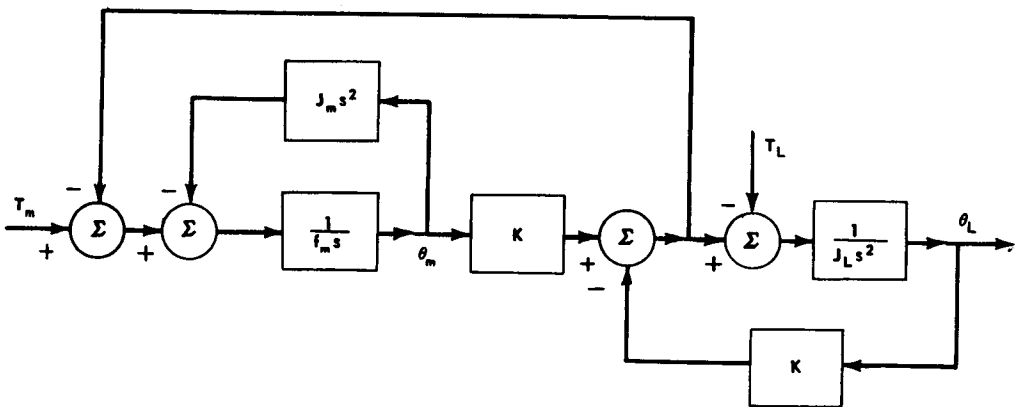
| RULE | ORIGINAL DIAGRAM | EQUIVALENT DIAGRAM |
|------|------------------|---|
| 11 | | |
| 12 | | <p style="text-align: center;">WHERE $\Delta_1 = AC - BD$</p> |
| 13 | | <p style="text-align: center;">WHERE $\Delta_2 = 1 - ABCD$</p> |

Fig. 3-6 Block diagram manipulation and reduction "rules". (Sheet 3 of 3)

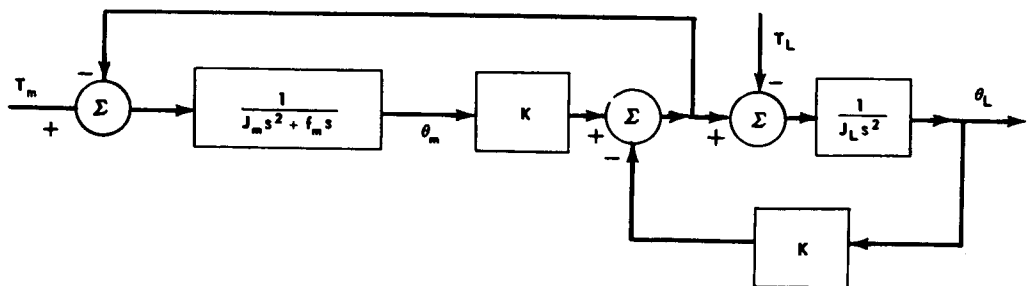
THEORY



A. ELEMENTARY BLOCK DIAGRAM



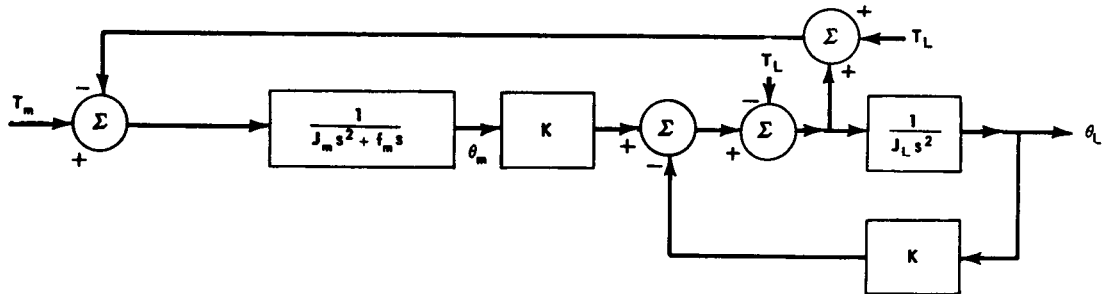
B. USE OF RULE 9 OF FIG. 3.6



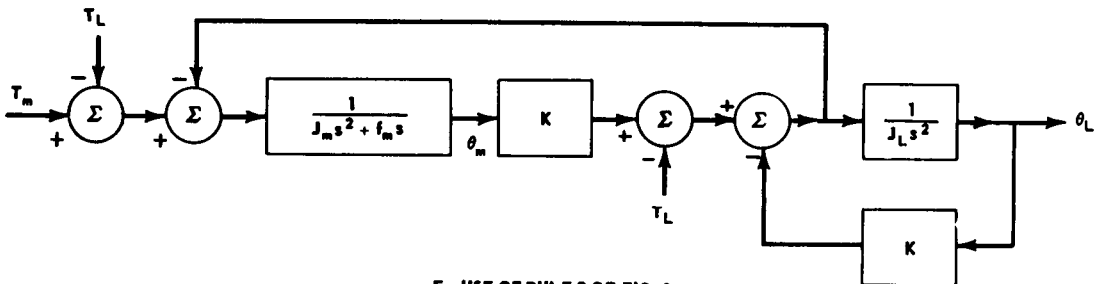
C. USE OF RULE 3 OF FIG. 3.6

Fig. 3.7 Block diagram examples. (Sheet 1 of 3)

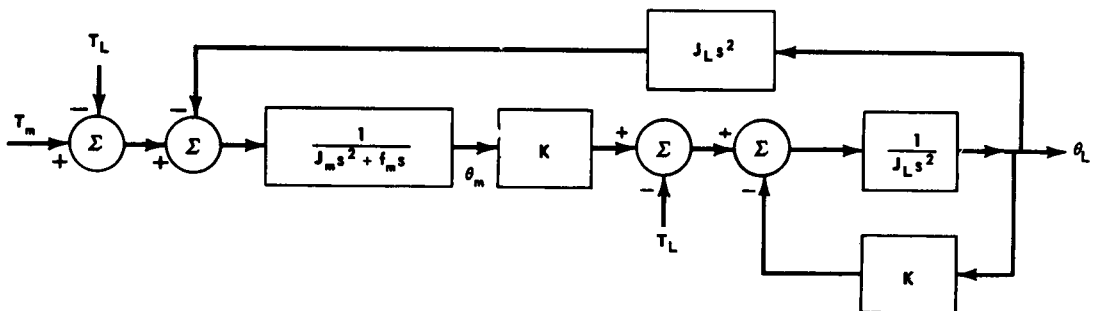
DETERMINING DYNAMIC RESPONSE OF LINEAR SYSTEMS



D. USE OF RULE 11 OF FIG. 3-6



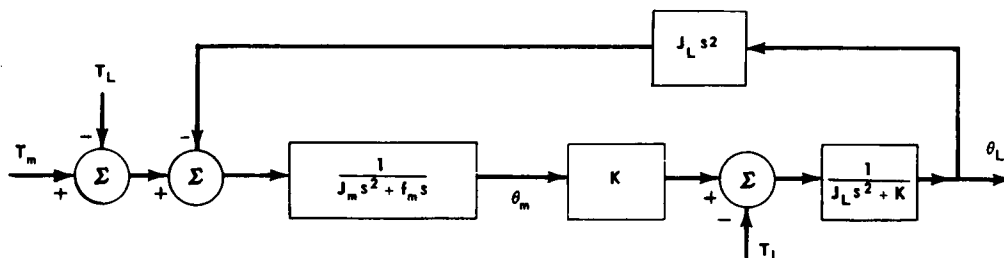
E. USE OF RULE 9 OF FIG. 3-6



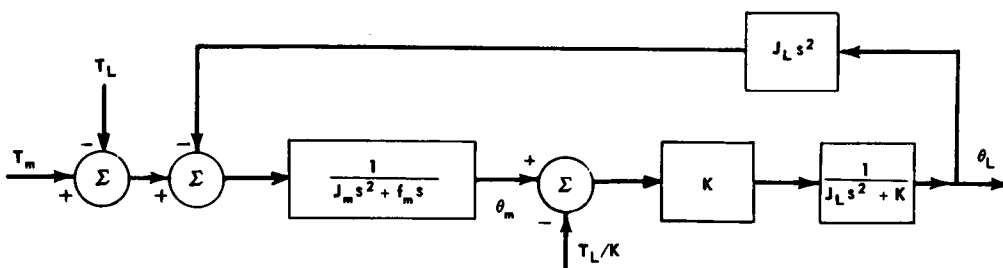
F. USE OF RULES 6 AND 8 OF FIG. 3-6

Fig. 3-7 Block diagram examples. (Sheet 2 of 3)

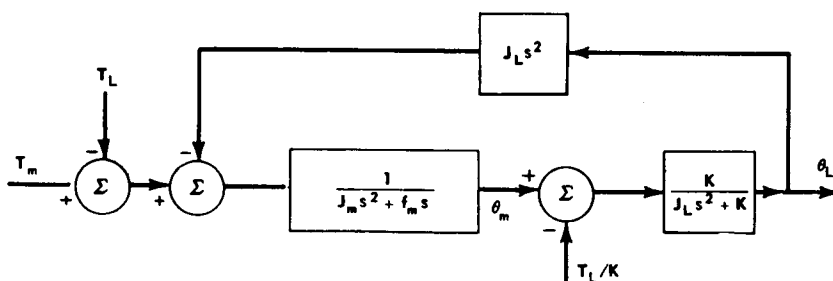
THEORY



G. USE OF RULE 3 OF FIG. 3-6



H. USE OF RULE 1 OF FIG. 3-6



I. USE OF RULE 2 OF FIG. 3-6

Fig. 3-7 Block diagram examples. (Sheet 3 of 3)

CHAPTER 4

STABILITY OF FEEDBACK CONTROL SYSTEMS*

4-1 INTRODUCTION

The determination of system stability is the first step in the design of any linear control system. To carry out this first step, a test for system stability is required. The particular stability test used will depend on the meaning attached to the term stable operation. Generally, a system is said to be stable if it remains at rest when all inputs are zero and if (for any disturbance) no signal grows without bound or exhibits sustained oscillation when the inputs are returned to zero. In the case of linear systems, the only situation in which unstable behavior can occur is the one in which the roots of the characteristic equation of the closed-loop system lie in the right-half s plane and therefore have positive real parts. The response modes corresponding to right-half-plane roots of the characteristic equation have amplitudes that increase without limit as time increases. Consequently, any stability criterion for a linear system is essentially a method of determining whether or not the characteristic equation has right-half-plane roots.

In Fig. 4-1 the general single-loop system is shown. The output response transform for this system is

$$C(s) = \frac{G_1(s) G_2(s) R(s) - G_2(s) U(s)}{1 + G_1(s) G_2(s) H(s)} \quad (4-1)$$

The characteristic equation of the system is

$$1 + G_1(s) G_2(s) H(s) = 0 \quad (4-2)$$

Thus, if any of the roots of Eq. (4-2) lie in the right-half s plane, the system of Fig. 4-1 is unstable.

The presence of right-half-plane roots of the denominator of the response transform $C(s)$ [i.e., right-half-plane poles of $C(s)$] could be determined by direct factorization of Eq. (4-2), after it has been cleared of fractions. Since a system is unstable if *one or more* right-half-plane poles of $C(s)$ exist, it is usually sufficient to determine whether these poles exist; however, it is not necessary to determine their exact location. Hence, the standard stability criteria that are discussed in this chapter (with the exception of the root-locus method) merely determine the number of unstable poles without regard to their location in the right-half s plane.

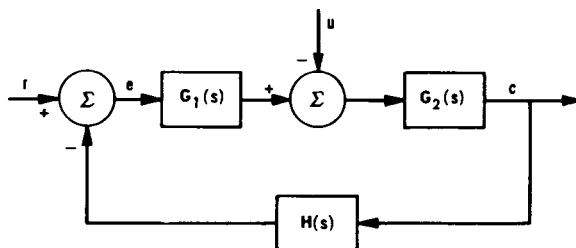


Fig. 4-1 Single-loop system — block diagram.

*By L. A. Gould

Three useful stability criteria — the Routh criterion, the Nyquist criterion, and root-locus method — are described in succeeding paragraphs.

The Routh criterion is the simplest to apply and can be used when the characteristic equation is known, at least in literal form. This criterion can be employed to determine the presence of any roots of an algebraic equation that lie in the right-half plane. If the equation is the characteristic equation of a closed-loop system, the presence of right-half-plane roots means that the system is unstable.

The Nyquist criterion is the most widely used stability criterion because the only information it requires for its application is a plot of the open-loop frequency response $G_1(j\omega) G_2(j\omega) H(j\omega)$. This frequency response can be determined either from an analytical representation of component behavior or from direct measurement of the response of the components to sinusoidal inputs. In addition to its use in determining the presence of system stability, the Nyquist criterion is extended in many design procedures to give an indication of the *degree of stability* possessed by the stable system (see Chs. 5 and 6). By an examination of the behavior of the $G_1(j\omega) G_2(j\omega) H(j\omega)$ locus in the vicinity

of the $-1 + j0$ point, the Nyquist criterion provides the servo engineer with a relatively straightforward and extremely powerful tool for analysis and design.

The root-locus method is a graphical technique for revealing the position of the poles of the response transform $C(s)$ in the s plane as a gain factor of the open-loop function $G_1(s) G_2(s) H(s)$ is varied. The primary advantage of this method for stability determination is that the closed-loop pole locations are kept in evidence at all times. Thus, it is easy to see when the poles move into the right-half plane as the gain factor is varied. There are two primary disadvantages connected with the root-locus method. First, the location of the poles *and* zeros of the open-loop function must be specified. This often requires some sort of analytical approximation to the experimental test data. Second, the plotting of the paths of the closed-loop poles involves a trial-and-error procedure that can be quite tedious. In spite of these disadvantages, however, the root-locus method is quite useful in that it immediately places in evidence the closed-loop pole-zero configuration for any particular design (stable, of course). Thus, the characteristics of the time response of the system are easily ascertained and the verification of performance specifications in the time domain is a straightforward matter.

4-2 ROUTH CRITERION^(1,2,6)

By applying the Routh stability criterion, one can determine whether any roots of an algebraic equation lie in the right-half s plane. If the coefficients of the equation are known only in literal form, the Routh criterion yields only a set of inequality conditions for stability. However, if the coefficients of the equation are known numerically, the criterion permits one to determine whether stability actually exists.

To show the general procedure used in applying the Routh criterion, consider the following general algebraic equation:

$$a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0 = 0 \quad (4-3)$$

Next, the coefficients are arrayed in two rows, alternate coefficients being placed in alternate rows

$$\begin{array}{cccccc} (1) & a_n & a_{n-2} & a_{n-4} & \dots & 0 \\ (2) & a_{n-1} & a_{n-3} & a_{n-5} & \dots & 0 \end{array} \quad (4-4)$$

Then, the array is extended by taking appropriate cross-products to determine the elements in the third row

$$(3) \quad \frac{a_{n-1} a_{n-2} - a_n a_{n-3}}{a_{n-1}} \quad \frac{a_{n-1} a_{n-4} - a_n a_{n-5}}{a_{n-1}} \dots \quad (4-5)$$

STABILITY OF FEEDBACK CONTROL SYSTEMS

The elements of the fourth row are formed by taking cross-products of the elements of the second and third rows, in exactly the same manner that the third-row elements were formed. This process is continued until all the elements of a row are zero. On completion of the array, the Routh criterion can be employed to determine the presence of right-half-plane roots.

The Routh criterion states:

The number of roots of the original equation that lie in the right-half s plane equals the number of sign

changes in the elements that form the first column of the final array.

An examination of the procedure used above shows that all the elements of any row after the second may be divided by a positive number without changing the result.

Examples.

(a) Consider the following algebraic equation with lateral coefficients:

$$s^4 + Ks^3 + 2s^2 + 4s + M = 0 \quad (4-6)$$

where $K > 0$ and $M > 0$.

The complete algebraic array is as follows:

| | | | | | |
|-----|------------------------------|------|---|---|--------------------------|
| (1) | 1 | 2 | M | 0 | |
| (2) | K | 4 | 0 | 0 | |
| (3) | (K-2) | KM/2 | 0 | 0 | (row multiplied by K/2) |
| (4) | [(K-2) - K ² M/8] | 0 | 0 | 0 | (row multiplied by 1/4) |
| (5) | (K-2-K ² M/8) | 0 | 0 | 0 | (row multiplied by 2/KM) |
| (6) | 0 | 0 | 0 | 0 | |

The inequalities that determine stability are

$$K > 0, \quad (4-8)$$

$$K-2 > 0, \quad (4-9)$$

and

$$[(K-2) - K^2M/8] > 0 \quad (4-10)$$

(b) As an example of an algebraic equation with numerical coefficients, consider the equation

$$s^5 + 2s^4 + 2s^3 + 46s^2 + 89s + 260 = 0 \quad (4-11)$$

The complete numerical array is as follows:

THEORY

| | | | | | |
|-----|----|-------|-----|-------------------------|--------|
| (1) | 1 | 2 | 89 | 0 | |
| (2) | 2 | 46 | 260 | 0 | |
| (3) | -1 | -1.95 | 0 | 0 (row divided by 21) | (4-12) |
| (4) | 1 | 6.17 | 0 | 0 (row divided by 42.1) | |
| (5) | 1 | 0 | 0 | 0 (row divided by 4.22) | |
| (6) | 1 | 0 | 0 | 0 (row divided by 6.17) | |
| (7) | 0 | 0 | 0 | 0 | |

Inspection of the signs of the elements of the first column shows that one sign change occurs in going from the second to the third row and another in going from the third to the fourth row. Hence, two roots of Eq. (4-11) lie in the right-half s plane. The factors of Eq. (4-11) are

$$(s + 4)(s - 2 + j3)(s - 2 - j3)$$

$$(s + 1 + j2)(s + 1 - j2) \quad (4-13)$$

The right-half s -plane roots of Eq. (4-13) are

$$s_1, s_2 = +2 \pm j3 \quad (4-14)$$

4-3 NYQUIST CRITERION^(3,6)

The Nyquist criterion is a graphical procedure by which one can determine whether any of the roots of the equation

$$1 + G(s) = 0 \quad (4-15)$$

lie in the right-half s plane. Only the following information is required in this procedure: (1) the magnitude and phase angle of $G(j\omega)$; (2) the behavior of $G(s)$ at the poles of $G(s)$ that lie on the imaginary axis or at the origin of the s plane; and (3) the number of poles of $G(s)$ in the right-half s plane. (NOTE: For nonunity feedback loops, one tests for the zeros of the function $1 + G(s)H(s)$ where $G(s)$ is the forward transfer function and $H(s)$ is the feedback transfer function.)

The Nyquist criterion can be expressed mathematically as

$$Z = N + P \quad (4-16)$$

where

Z = number of zeros of $1 + G(s)$ that lie in the right-half s plane

P = number of poles of $G(s)$ that lie in the right-half s plane

N = number of clockwise encirclements of the point $-1 + j0$ by the locus of $G(s)$ as s describes the path shown in Fig. 4-2

For stability, Z must be zero; that is, $P = -N$. If $P \neq -N$, the system is unstable.

If there are any poles of $G(s)$ on the imaginary axis, the $G(j\omega)$ locus will become infinite at these points. To determine the behavior of the $G(j\omega)$ locus at these poles, so as to be able to count encirclements, a small

STABILITY OF FEEDBACK CONTROL SYSTEMS

semicircular detour is made into the right-half s plane at each pole of $G(s)$ on the imaginary axis. Thus, the $G(j\omega)$ locus will describe a large semicircle instead of becoming infinite. If the pole on the imaginary axis is of multiple order, the $G(j\omega)$ locus will describe one semicircle for each order of the multiple pole.

A convenient rule for determining the direction of turn of the $G(j\omega)$ locus at the imaginary-axis poles of $G(s)$ is

Turn to the *right* by 180° for each order of the pole as the frequency increases.

If $G(s)$ has no poles in the right-half s plane or on the imaginary axis (except at the origin), the Nyquist stability criterion simplifies to†

$$|G(j\omega)| < 1 \quad \text{when } \angle G(j\omega) = -180^\circ \quad (4-17)$$

Examples.

(a) Consider the function

$$G(s) = \frac{K}{s(Ts + 1)^2} \quad (4-18)$$

For what range of K will $1 + G(s)$ have stable roots?

†Symbol \angle denotes "angle"

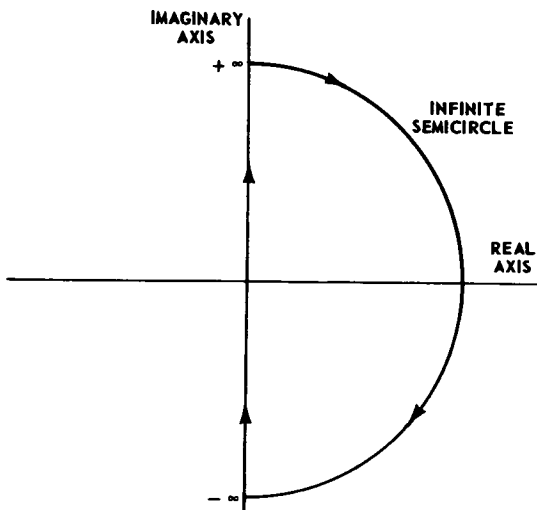


Fig. 4-2 Locus of s for the Nyquist criterion.

To simplify the calculation, change variables, letting $\lambda = Ts$. Then,

$$G(\lambda) = \frac{KT}{\lambda(\lambda + 1)^2} \quad (4-19)$$

Plot $(1/KT)G(\lambda)$ on the complex plane for $\lambda = j\omega$. Such a plot is sketched in Fig. 4-3. Since encirclements depend *only* upon the topology of the plot, the locus can be distorted to facilitate the counting of encirclements (Fig. 4-4).

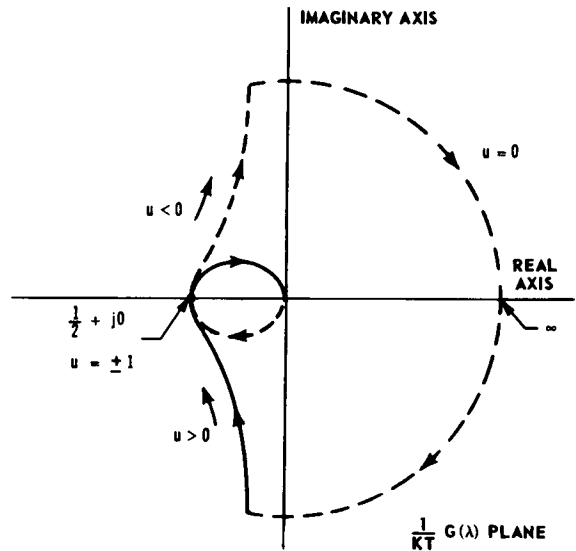


Fig. 4-3 Locus of $\frac{1}{\lambda(\lambda + 1)^2}$ for $\lambda = j\mu$.

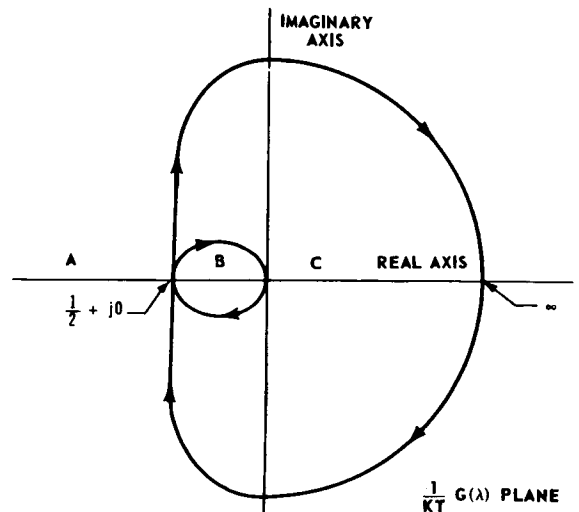


Fig. 4-4 Distortion of locus of $\frac{1}{\lambda(\lambda + 1)^2}$.

THEORY

It is convenient to arrange the pertinent information in tabular form as follows:

| Location of $-1 + j0$ | P | N | Z | Nature of Stability |
|--------------------------|-----|-----|-----|-------------------------|
| A | 0 | 0 | 0 | stable ($Z = 0$) |
| B | 0 | 2 | 2 | unstable ($Z \neq 0$) |
| C | 0 | 1 | 1 | unstable ($Z \neq 0$) |

The stable range of KT is determined from the table above and Eq. (4-17) as

$$0 < KT < 2.0$$

This expression is found by determining the range of KT which, when multiplying the $(1/KT)G(\lambda)$ locus, will keep the point $-1 + j0$ in region A .

(b) Consider the function

$$G(s) = K \frac{(1+s)}{s(1-s)} \quad (4-20)$$

For what range of K will $1 + G(s)$ have stable roots?

The $(1/K)G(s)$ locus is sketched in Fig. 4-5. The distorted locus for counting encirclements appears in Fig. 4-6. A table of the pertinent information is given below:

| Location of $-1 + j0$ | P | N | Z | Nature of Stability |
|--------------------------|-----|-----|-----|------------------------|
| A | 1 | 0 | 1 | unstable |
| B | 1 | -1 | 0 | stable |
| C | 1 | 1 | 2 | unstable |

Stable roots exist for K in the range

$$-\infty < K < -1$$

These limits are found by determining the range of K which, when multiplying the $(1/K)G(s)$ locus, will keep the $-1 + j0$ point in region B .

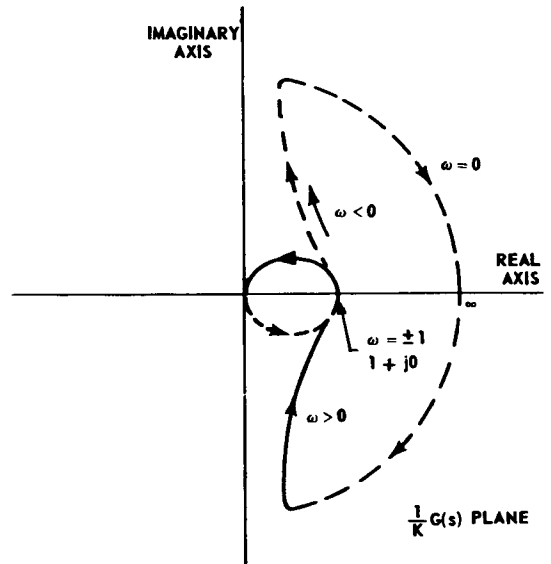


Fig. 4-5 Locus of $\frac{1}{s} \left(\frac{1+s}{1-s} \right)$.

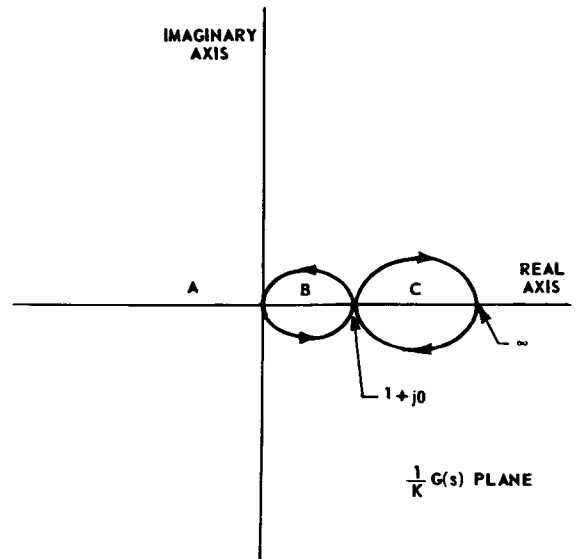


Fig. 4-6 Distortion of locus of $\frac{1}{s} \left(\frac{1+s}{1-s} \right)$.

4-4 ROOT-LOCUS METHOD^(4,5)

In the root-locus method, a plot is made of the locus of the roots of $1 + G(s) = 0$ as a function of a gain factor of $G(s)$. $G(s)$ must be known numerically in completely factored form. For nonunity feedback loops, one plots the roots of the equation $1 + G(s) H(s) = 0$ where $G(s)$ is the forward transfer function and $H(s)$ is the feedback transfer function.

If $G(s)$ is written in the form

$$G(s) = K_o \frac{\prod_{k=1}^z (s + s_k)}{\prod_{j=1}^p (s + s_j)} \quad (4-21)$$

where K_o is varied from 0 to $+\infty$, then the necessary condition for a point in the s plane to lie on the locus of the roots of $1 + G(s) = 0$, as K_o varies, is

$$\sum A_z - \sum A_p = -180^\circ \quad (4-22)$$

where

$\sum A_z$ = sum of the angles of the phasors from the *zeros* of $G(s)$ to the point in question

$\sum A_p$ = sum of the angles of the phasors from the *poles* of $G(s)$ to the point in question

The value of the constant K_o that is associated with each root-locus point is found from the relation

$$K_o = \frac{\prod |V_p|}{\prod |V_z|} \quad (4-23)$$

where

$\prod |V_p|$ = product of the magnitudes of the phasors from the *poles* of $G(s)$ to the root-locus point

$\prod |V_z|$ = product of the magnitudes of the phasors from the *zeros* of $G(s)$ to the root-locus point

The root-locus method can be used to reveal the position of the roots of $1 + G(s) = 0$ directly and to determine whether any of the

roots can move into the right-half s plane as the constant K_o is varied. This method of stability determination is primarily a graphical one, particularly when determining the points in the s plane that satisfy the angle condition [Eq. (4-22)]. Although the angle condition determines the entire locus, it is still necessary to find the actual points by a trial-and-error procedure. That is, a point is guessed and the angle condition is checked; if the angle condition is not satisfied, another point is tried, etc.

To facilitate the plotting of the root locus, several theorems based on the angle condition [Eq. (4-22)] and the magnitude condition [Eq. (4-23)] have been established. These are:

(a) The number of branches for a given locus equals the number of roots of $1 + G(s) = 0$.

(b) The locus starts ($K_o = 0$) at poles and ends ($K_o = \infty$) at zeros.

(c) The real-axis position of the locus always has an odd number of poles and zeros to the right of the s point for $K_o > 0$.

(d) The breakaway from the real axis into the complex plane between two adjacent poles occurs at the point of maximum K_o .

(e) For two adjacent zeros, the locus enters the real axis from the complex plane at the point of minimum K_o .

(f) Near complex *poles*, the direction of the locus is given by

$$[180^\circ - \sum A_z + \sum A_p]$$

where $\sum A_z$ is the sum of the angles of the phasors from all the other *zeros* to the complex pole in question, and $\sum A_p$ is the sum of the angles of the phasors from all the other *poles* to the complex pole in question. Near complex *zeros*, the direction of the locus is given by

$$[-180^\circ + \sum A_p - \sum A_z].$$

THEORY

(g) The asymptotes of the locus for large values of s are given by a set of straight lines that intersect the real axis at angles

$$A = \frac{180^\circ \pm 360^\circ n}{p - z} \quad (n = 0, 1, 2, \dots) \quad (4-24)$$

and whose intersection with the real axis is given by the centroid of the pole-zero configuration

$$x_o = \frac{\sum_{j=1}^p s_j - \sum_{k=1}^z s_k}{p - z} \quad (4-25)$$

where

$$\begin{aligned} p &= \text{number of poles} & s_j &= j^{\text{th}} \text{ pole} \\ z &= \text{number of zeros} & s_k &= k^{\text{th}} \text{ zero} \end{aligned}$$

(h) The locus is symmetrical with respect to the real axis.

These theorems may be verified for the various loci in Fig. 4-7, which presents examples of a wide variety of root loci for systems up to the fourth order. In this figure, $T(p, z)$ indicates a system with p poles and z zeros.

As an example of a typical root-locus plot for a unity-feedback loop, consider the function

$$\begin{aligned} G(s) &= \frac{K}{s(s/\omega_1 + 1)(s/\omega_2 + 1)} \\ &= \frac{K\omega_1\omega_2}{s(s + \omega_1)(s + \omega_2)} \end{aligned}$$

where $\omega_1 = 10$ rad/sec, $\omega_2 = 30$ rad/sec, and the conditions are

$$\frac{s}{s} + \frac{s}{s + 10} + \frac{s}{s + 30} = 180^\circ \quad (\text{angle condition})$$

$$|s| |s + 10| |s + 30| = 300 K \quad (\text{magnitude condition})$$

The location of the poles of $G(s)$ on the s plane is as shown in Fig. 4-8.

The root locus coincides with the real axis lying between the pole at the origin and the pole at -10 , as well as with the part of the real axis lying to the left of the pole at -30 . The locus breaks away from the real axis at

some point between the pole at the origin and the pole at -10 . To locate this breakaway point, either the technique described in theorem (d) or, in this simple case, an analytical technique can be applied.

Let $-\delta$ be the location of the breakaway point. With $s = -\delta$, the magnitude condition becomes

$$f(\delta) = \delta(-\delta + 10)(-\delta + 30) = 300 K$$

The negative of the value of δ which maximizes the left side of the equation above is the coordinate of the breakaway point. Thus

$$f(\delta) = \delta^3 - 40\delta^2 + 300\delta$$

$$f'(\delta) = 3\delta^2 - 80\delta + 300 = 0$$

$$\delta = 20.8, 6.3$$

Only the smaller value of δ satisfies the angle condition on the real axis. Therefore, the breakaway point is at -6.3 . The gain factor K at this point is obtained by substituting this value of δ into the magnitude condition equation, i.e.,

$$300 K = 6.3(-6.3 + 10)(-6.3 + 30)$$

$$K = 1.84$$

The point on the branch of the locus to the left of the pole at -30 , for this same value of gain K , is at -30.8 .

The asymptotes of the locus, for large values of s , intersect the real axis at angles given by the relation

$$A = \frac{180^\circ \pm 360^\circ n}{p - z}$$

Since $p = 3$ and $z = 0$, the angles are 60° and -60° (or $+120^\circ$). The real-axis intercept of the asymptotes is given by

$$\begin{aligned} x_o &= \frac{\sum s_j - \sum s_k}{p - z} = \frac{[0 + 10 + 30] - [0]}{3 - 0} \\ &= 13.3 \end{aligned}$$

A sketch of the data obtained so far is given in Fig. 4-9.

At this point, it would seem necessary to apply the exploratory s-point method to determine the rest of the locus. However, even before this is done, we can determine which of the asymptotes the locus approaches. In this particular case, we already know (from a Nyquist plot) that the function $1 + G(s)$ will have right-half-plane roots when the sensitivity is above a certain value. Therefore, the two branches of the locus in the complex portion of the plane must head toward the right-half plane as the gain factor K increases. Thus, the locus will cross the imaginary axis and go into the right-half plane. The points of imaginary-axis crossing are easy to find in this particular case because of the small number of poles involved in the configuration.

Consider the geometric properties of Fig. 4-10. At the crossing point, the angle condition requires that

$$\tan^{-1} \frac{\omega_c}{\omega_1} + \tan^{-1} \frac{\omega_c}{\omega_2} + 90^\circ = 180^\circ$$

Therefore

$$\frac{\left(\frac{\omega_c}{\omega_1}\right) + \left(\frac{\omega_c}{\omega_2}\right)}{1 - \left(\frac{\omega_c^2}{\omega_1 \omega_2}\right)} \rightarrow \infty$$

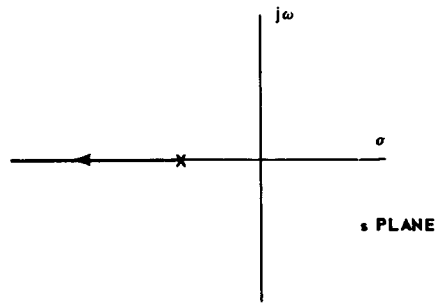
or

$$\omega_c = \sqrt{\omega_1 \omega_2} = \sqrt{300} = 17.3$$

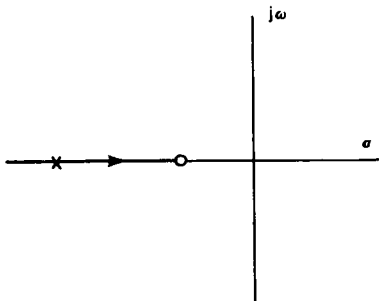
At this point, the gain K is determined by the magnitude condition. Using this condition, it is found that $K = 40$. The corresponding point on the branch to the left of the pole at -30 is at -40 .

The remainder of the locus can be sketched in, or a more accurate determination of the locus points can be made by the exploratory s-point method. Figure 4-11 is a sketch of the entire locus, with key points indicated.

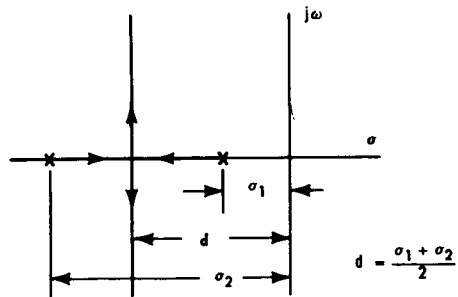
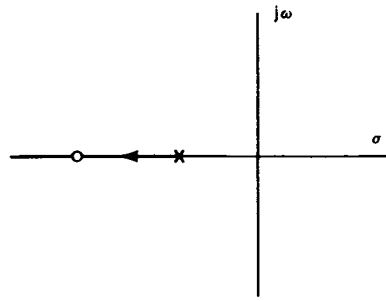
THEORY



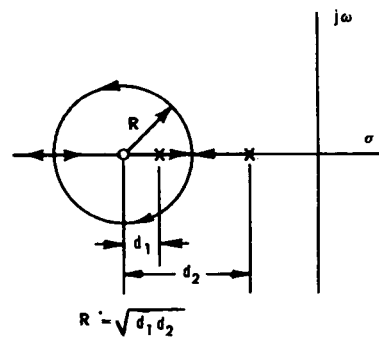
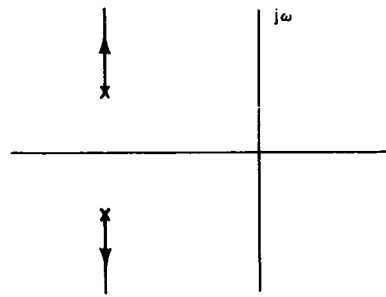
A. TYPE T(1,0)



B. TYPE T(1,1)



C. TYPE T(2,0)



D. TYPE T(2,1)

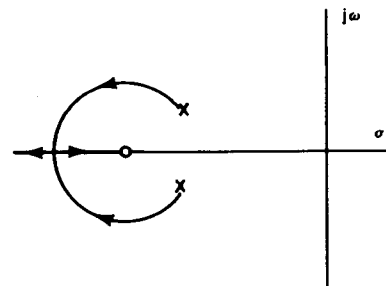


Fig. 4-7 Root-loci plots. (Sheet 1 of 5)

Adapted from "The Study of Transients in Linear Feedback Systems by Conformal Mapping and the Generalized Root Locus Method", by V. C. M. Yeh, ScD Thesis M.E., 1952, Massachusetts Institute of Technology.

STABILITY OF FEEDBACK CONTROL SYSTEMS

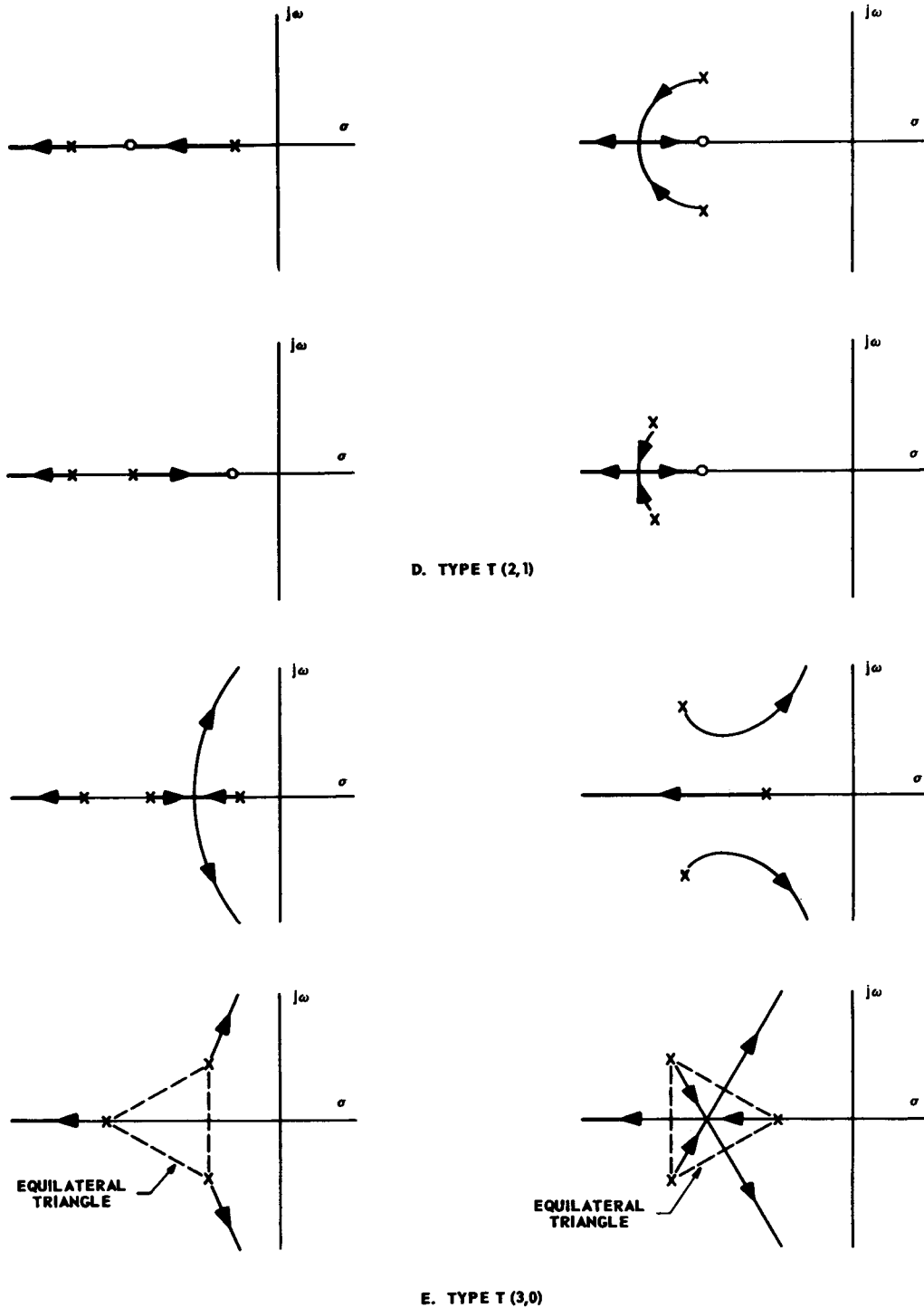
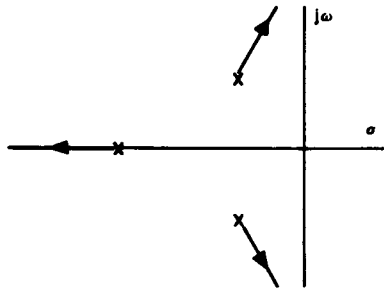
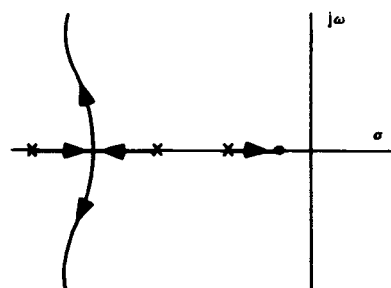
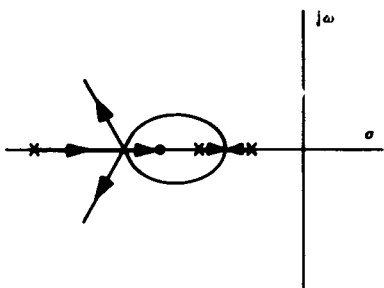
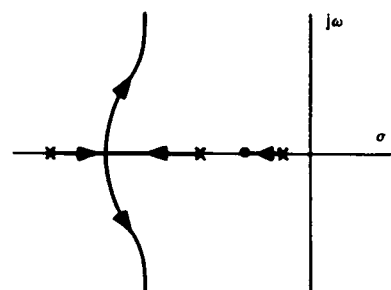
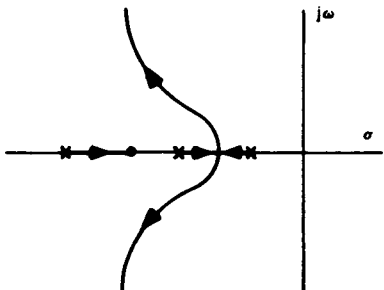
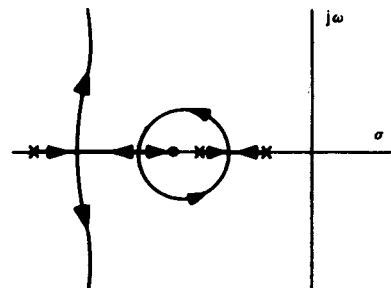
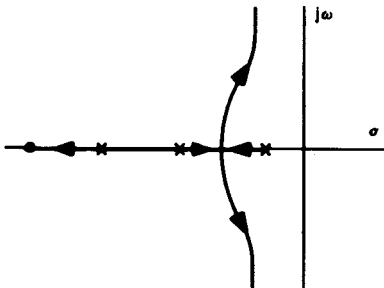
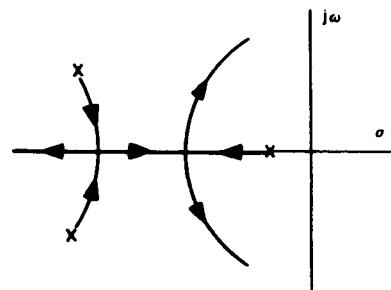


Fig. 4-7 Root-loci plots. (Sheet 2 of 5)

THEORY



E. TYPE T(3,0)



F. TYPE T(3,1)

Fig. 4-7 Root-loci plots. (Sheet 3 of 5)

STABILITY OF FEEDBACK CONTROL SYSTEMS

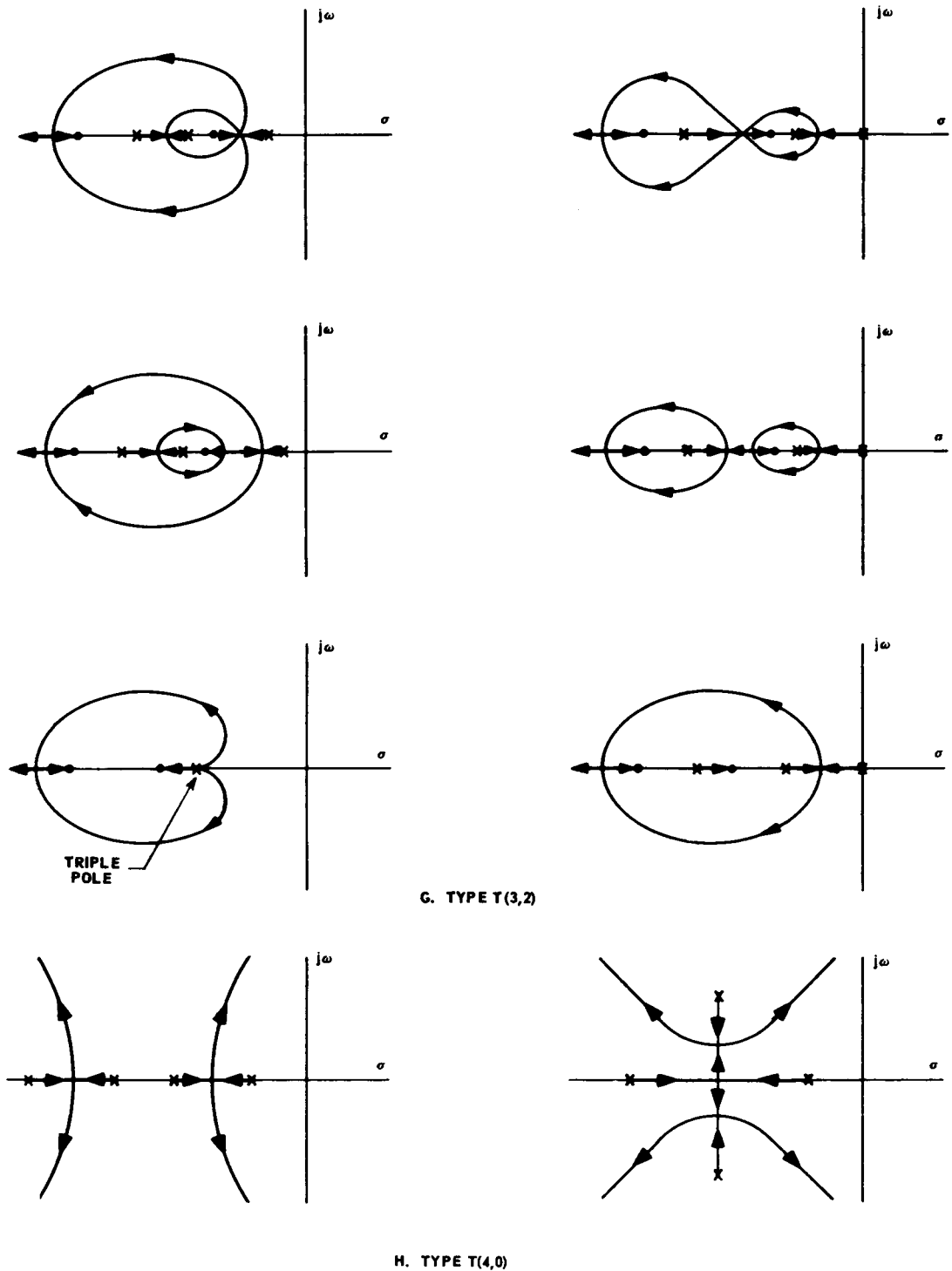


Fig. 4-7 Root-loci plots. (Sheet 4 of 5)

THEORY

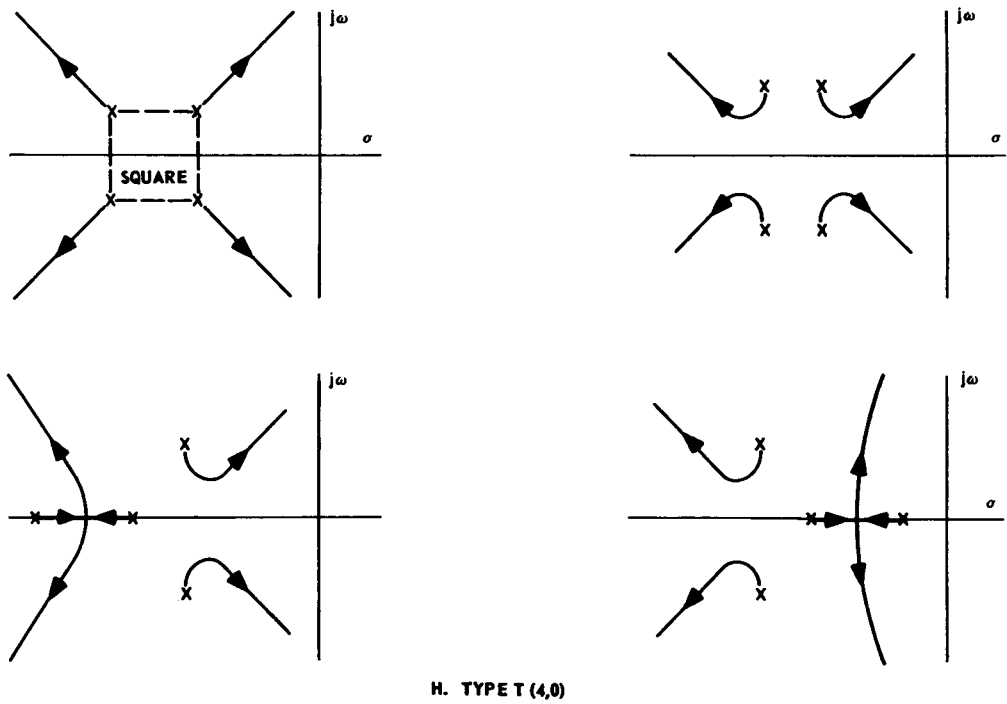


Fig. 4-7 Root-loci plots. (Sheet 5 of 5)

STABILITY OF FEEDBACK CONTROL SYSTEMS

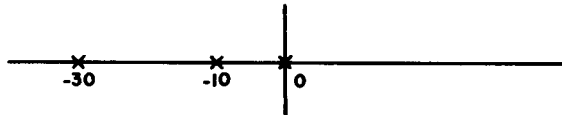


Fig. 4-8 Poles for $G(s) = \frac{300K}{s(s+10)(s+30)}$.

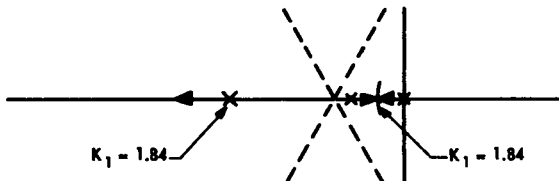


Fig. 4-9 Asymptotes and real axis behavior for

$$G(s) = \frac{300K}{s(s+10)(s+30)}.$$

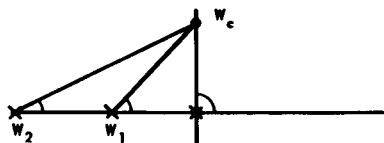


Fig. 4-10 Construction for determining imaginary

axis crossing $G(s) = \frac{300K}{s(s+10)(s+30)}$
when $0 < K < \infty$.

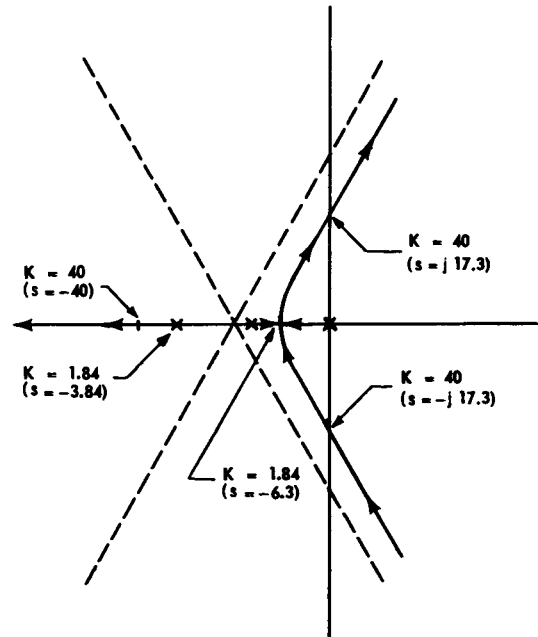


Fig. 4-11 Complete root locus for $G(s)$

$$= \frac{300K}{s(s+10)(s+30)} \text{ when } 0 < K < \infty.$$

THEORY

BIBLIOGRAPHY

- 1 M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, Vol. I, pp. #197-201, John Wiley & Sons, Inc., New York, N. Y., 1942.
- 2 W. R. Ahrendt and J. F. Taplin, *Automatic Feedback Control*, pp. #78-80, McGraw-Hill Book Company, Inc., New York, N. Y., 1951.
- 3 H. W. Bode, *Network Analysis and Feedback Amplifier Design*, pp. #137-169, D. Van Nostrand Company, Inc., New York, N. Y., 1945.
- 4 W. R. Evans, *Control System Dynamics*, pp. #96-121, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.
- 5 W. R. Evans, "Control System Synthesis by Root-Locus Method", *Trans. AIEE*, Vol. 69, Part I, pp. #66-69, 1950.
- 6 E. A. Guillemin, *The Mathematics of Circuit Analysis*, pp. #395-409, John Wiley & Sons, Inc., New York, N. Y., 1949.

CHAPTER 5

GAIN DETERMINATION*

5-1 PERFORMANCE CRITERIA AND DEFINITIONS^(3,6)

5-1.1 GENERAL

Performance criteria are tests or rules by which one can determine, from the system parameters, whether or not the system has certain particular performance characteristics. An important parameter used in the design of servo systems is the gain of the system. The relations between gain and performance specifications are usually expressed in terms of performance constants, such as the velocity constant, the acceleration constant, and the torque constant. Definitions of several important parameters and constants are given below.

5-1.2 GAIN

If $G(s)$ is the transfer function of a component, the gain K of the component is

$$K = \lim_{s \rightarrow 0} s^{\pm n} G(s) \quad (5-1)$$

where n is the order of the pole or zero of $G(s)$ at the origin. The plus sign is used for a simple or multiple *pole* at the origin, and the minus sign is used for a simple or multiple *zero* at the origin. The gain K is easy to identify if $G(s)$ is written in the form

$$G(s) = K s^{\pm n} \left[\frac{a_k s^k + a_{k-1} s^{k-1} + \dots + 1}{b_j s^j + b_{j-1} s^{j-1} + \dots + 1} \right] \quad (5-2)$$

5-1.3 VELOCITY CONSTANT

The velocity constant of a system is a measure of the steady-state error if the input to the system is a constant velocity. The velocity constant is defined by the relation

$$K_v = \left(\frac{E_{ss}}{\omega_i} \right)^{-1} \quad (5-3)$$

where

ω_i = constant input velocity

E_{ss} = steady-state error

For a single-loop unity-feedback system (Fig. 5-1), the velocity constant is

$$K_v = \lim_{s \rightarrow 0} \left[s \frac{C(s)}{E(s)} \right] = \lim_{s \rightarrow 0} [s G(s)] \quad (5-4)$$

An analysis of Eq. (5-4) shows that the velocity constant of a single-loop unity-feedback system is finite and nonzero only if the open-loop transfer function $C(s)/E(s)$ has exactly one *single-order pole* at the origin (one integration).

5-1.4 ACCELERATION CONSTANT

The acceleration constant of a system is defined by the relation

$$K_a = \left(\frac{E_{ss}}{\alpha_i} \right)^{-1} \quad (5-5)$$

where

α_i = constant input acceleration

E_{ss} = steady-state error

*By L. A. Gould

THEORY

For a single-loop unity-feedback system (Fig. 5-1), the acceleration constant is

$$K_A = \lim_{s \rightarrow 0} \left[s^2 \frac{C(s)}{E(s)} \right] = \lim_{s \rightarrow 0} [s^2 G(s)] \quad (5-6)$$

An analysis of Eq. (5-6) shows that the acceleration constant of a single-loop unity-feedback system is finite and nonzero only if the open-loop transfer function $C(s)/E(s)$ has exactly one *double-order* pole at the origin (two integrations).

5-1.5 TORQUE CONSTANT

The torque constant of a system is defined by the relation

$$K_T = \left(\frac{E_{ss}}{T_L} \right)^{-1} \quad (5-7)$$

where

T_L = constant load torque

E_{ss} = steady-state error

5-1.6 STATIC ACCURACY

The static accuracy of a *linear* system is measured by the steady-state error that is developed for a specified steady-state input or disturbance. A steady-state input in this context means that the input is a constant position, a constant velocity, a constant acceleration, etc. A steady-state disturbance means that the disturbance is a constant. The performance constants defined previously will uniquely determine these steady-state errors, as may be seen from the definitions of the constants.

5-1.7 BANDWIDTH

In general, the bandwidth of a servo system refers to a frequency interval between 0 and some upper frequency. There is no universally accepted definition of the upper frequency.

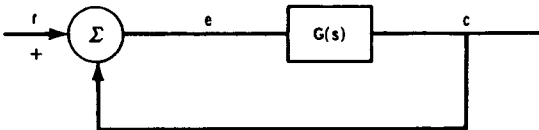


Fig. 5-1 Single-loop unity-feedback system.

Several commonly used upper-frequency values for unity-feedback systems (Fig. 5-1) are given below:

(a) ω_R , resonant frequency — frequency at which the closed-loop frequency response $C(j\omega)/R(j\omega)$ has its peak magnitude M_p (Fig. 5-2).

(b) ω_a — frequency at which the magnitude of the closed-loop frequency response $C(j\omega)/R(j\omega)$ is unity (Fig. 5-2).

(c) ω_b — frequency at which the magnitude of the closed-loop frequency response $C(j\omega)/R(j\omega)$ is 0.707 (Fig. 5-2).

(d) ω_p — frequency at which the phase of the closed-loop frequency response is -90° (Fig. 5-3).

(e) ω_e — frequency at which the magnitude of the error-to-input frequency response $E(j\omega)/R(j\omega)$ is 0.1 (Fig. 5-4).

(f) ω_{cm} , magnitude crossover frequency — frequency at which the magnitude of the open-loop frequency response $C(j\omega)/E(j\omega)$ is unity (Fig. 5-5).

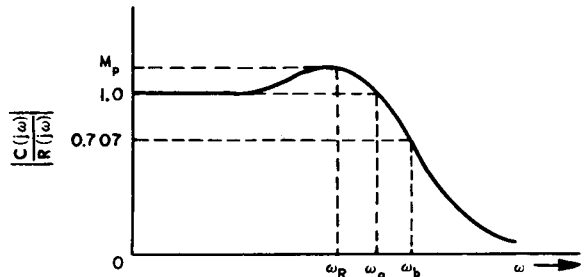


Fig. 5-2 Bandwidth measures from magnitude of closed-loop frequency response $C(j\omega)/R(j\omega)$.

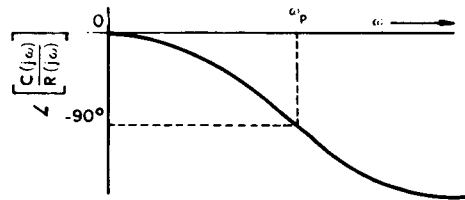


Fig. 5-3 Bandwidth measure from phase of closed-loop frequency response $C(j\omega)/R(j\omega)$.

GAIN DETERMINATION

(g) ω_c , asymptote crossover frequency — frequency at which the -10 dg/dec asymptote of the open-loop frequency response $C(j\omega)/E(j\omega)$ crosses 0 dg (Fig. 5-5; see Par. 5-3 for terminology).

5-1.8 PEAK MAGNITUDE

The peak magnitude M_p is defined as the maximum value of the magnitude of the closed-loop frequency response $C(j\omega)/R(j\omega)$ or the magnitude of the resonant peak of the response (see Fig. 5-2).

The value of M_p is used as a measure of the degree of stability, and design in the frequency domain usually involves adjusting a gain K so as to satisfy a specified value of M_p . Large values of M_p are indicative of highly oscillatory behavior, whereas values of M_p less than unity are indicative of heavily damped behavior. In practice, M_p usually lies between 1.3 and 1.6; that is, the range of M_p is usually specified as follows:

$$1.3 < M_p < 1.6 \quad (5-8)$$

or

$$1 \text{ dg} < 10 \log_{10} M_p < 2 \text{ dg} \quad (5-9)$$

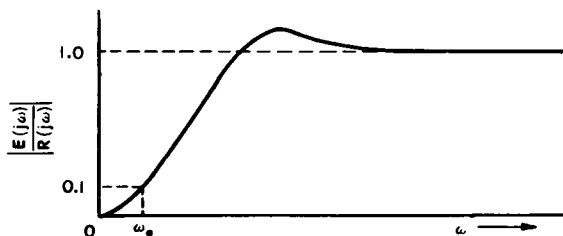


Fig. 5-4 Bandwidth measure from magnitude of error-to-input frequency response $E(j\omega)/R(j\omega)$.

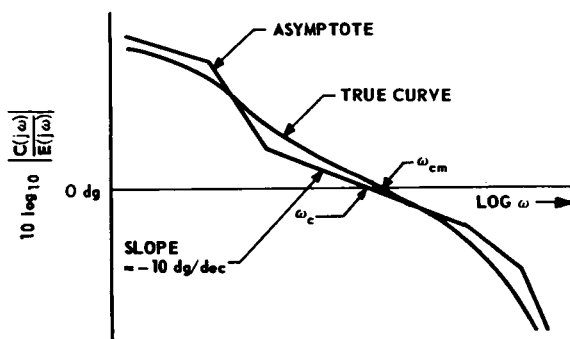


Fig. 5-5 Bandwidth measures from open-loop frequency response.

5-2 POLAR-PLANE REPRESENTATION ^(1,2,9,12)

5-2.1 GENERAL

A polar-plane representation of the open-loop frequency response $C(j\omega)/E(j\omega)$ is often used in the process of carrying out a design in the frequency domain. A plot of $C(j\omega)/E(j\omega)$ in polar coordinates makes it easier to apply the Nyquist stability criterion to determine gain setting ranges for stable operation. In addition, the determination of gain for a specified M_p value involves only a simple graphical construction on the polar plane (see Par. 5-4). Plots of both the direct function and its inverse are used, wherein only positive frequencies are usually considered.

5-2.2 DIRECT POLAR PLANE

A direct polar-plane plot of the function $G(j\omega)$ is constructed by drawing a curve through the points whose polar coordinates at each frequency are the magnitude of $G(j\omega)$ and the phase angle of $G(j\omega)$ at that frequency, where the phase angle of $G(j\omega)$ is the phase of $c(t)$ minus the phase of $e(t)$ when $e(t)$ and $c(t)$ are sinusoids. Positive angles are plotted in a counterclockwise direction. Increasing the gain associated with $G(j\omega)$ expands the polar locus in a radial direction. If $G(j\omega)$ is cascaded with another transfer function, the resulting polar coordinates of the combination are obtained at

THEORY

each frequency by: (1) multiplying the magnitude of $G(j\omega)$ by the magnitude of the cascaded function to give the magnitude of the combination; and (2) adding the phase angle of the cascaded function to the phase angle of $G(j\omega)$ to give the phase angle of the combination.

5-2.3 INVERSE POLAR PLANE

An inverse polar-plane plot of a function is a plot of the reciprocal of the function on the polar plane. The reciprocal or inverse of $G(j\omega)$ is written as follows:

$$G^{-1}(j\omega) = \frac{1}{G(j\omega)} \quad (5-10)$$

The polar coordinates of $G^{-1}(j\omega)$ at each frequency are given by: (1) the reciprocal of the magnitude of $G(j\omega)$; and (2) the negative of the phase angle of $G(j\omega)$. Increasing the gain of $G(j\omega)$ shrinks the inverse locus in the radial direction. If $G(j\omega)$ is cascaded with another transfer function, the resulting polar coordinates of the inverse of the combination are obtained at each frequency by: (1) multiplying the magnitude of the inverse $G(j\omega)$ locus by the inverse of the magnitude of the cascaded function to give the magnitude of the inverse of the combination; and (2) adding the negative of the phase angle of the cascaded function to the phase of the inverse $G(j\omega)$ locus to give the phase angle of the inverse of the combination.

Example. Plots of the direct function

$$G(j\omega) = \frac{1}{j\omega [(j\omega)^2 + 0.6j\omega + 1]} \quad (5-11)$$

and a multiple of its inverse $3G^{-1}(j\omega)$ appear in Fig. 5-6.

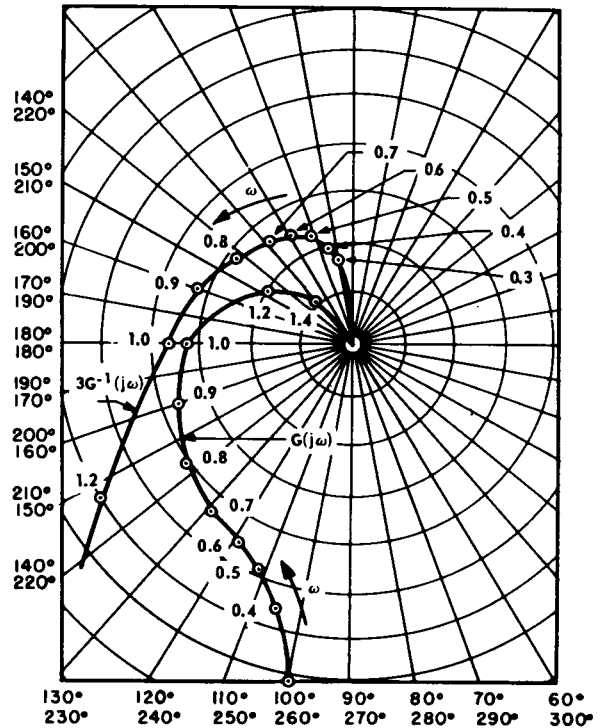


Fig. 5-6 Direct and inverse polar plots of

$$G(j\omega) = \frac{1}{j\omega [(j\omega)^2 + 0.6j\omega + 1]}$$

5-3 EXACT AND ASYMPTOTIC-LOGARITHMIC REPRESENTATIONS (5.9,13)

5-3.1 GENERAL

The logarithmic method of representing a function is a more convenient way to present frequency-response information than the polar-plane method. The advantage of the logarithmic procedure is that magnitude multiplication for cascaded functions reduces to the simple addition of logarithms. Further-

more, the magnitudes of the first- and second-order factors of transfer functions can readily be approximated by straight-line asymptotes when the functions are plotted to a logarithmic scale. Such asymptotic approximations reduce the time taken up by calculation and, in addition, enable the designer to make a rough estimate of system performance, when this is necessary.

GAIN DETERMINATION

5-3.2 SEPARATE MAGNITUDE AND PHASE PLOTS

The separate magnitude and phase-angle plots for a transfer function $G(j\omega)$ are respectively: (1) plots of $10 \log_{10} |G(j\omega)|$ versus $\log \omega$; and (2) plots of $\text{Ang } G(j\omega)$ versus $\log \omega$. The unit of logarithmic magnitude used in these plots is called the *decilog*, abbreviated *dg*. The magnitude of a number N in decilogs is $10 \log_{10} N$. For convenience in plotting, semilog graph paper is generally used. To plot a transfer function $G(j\omega)$ that is already in factored form, several aids (to be discussed below) are available which simplify the procedure. Before discussing these aids, however, it is helpful to point out the general types of factors that may appear in any rational algebraic function. Consider a function $G(j\omega)$ whose factored form can be written as follows:

$$G(j\omega) = K (j\omega)^{\pm n}$$

$$\frac{(T_1 j\omega + 1) \left[\left(j \frac{\omega}{\omega_{n_1}} \right)^2 + 2\zeta_1 j \frac{\omega}{\omega_{n_1}} + 1 \right] \dots}{(T_2 j\omega + 1) \left[\left(j \frac{\omega}{\omega_{n_2}} \right)^2 + 2\zeta_2 j \frac{\omega}{\omega_{n_2}} + 1 \right] \dots} \quad (5-12)$$

Only three general types of algebraic factors appear in Eq. (5-12).

The three factor types, which may occur in any rational function, are the following:

$$(j\omega)^{\pm n} \text{ (differentiation or integration)} \quad (5-13)$$

$$(Tj\omega + 1) \text{ (first order)} \quad (5-14)$$

$$\left[j \left(\frac{\omega}{\omega_n} \right)^2 + 2\zeta j \frac{\omega}{\omega_n} + 1 \right] \text{ (second order)} \quad (5-15)$$

5-3.3 MAGNITUDE CURVES

The magnitude curve of the quantity $(j\omega)^{\pm n}$ is a straight line passing through 0 dg at $\omega = 1$ with a slope equal to $\pm 10 n$ dg/decade.

The magnitude of the first-order factor $(Tj\omega + 1)^{\pm 1}$ can be approximated by two straight lines. For $T\omega \ll 1$, the asymptote is the 0-dg line. For $T\omega \gg 1$, the asymptote is a line with a slope of ± 10 dg/decade that crosses 0 dg at $T\omega = 1$. The frequency $\omega_b = 1/T$ is called the *break frequency* of the factor. The true magnitude curve can be obtained from the asymptotes by applying the two rules-of-thumb:

(a) At the break frequency, the true curve is 1.5 dg above (or below) the asymptotes.

(b) At an octave above and below the break frequency, the true curve lies 0.5 dg above (or below) the asymptotes.

The asymptotes and the true magnitude curves for a first-order factor are shown in Fig. 5-7. Note in this figure that the magnitude curve of $(Tj\omega = 1)^{-1}$ is the mirror image of the magnitude curve of $(Tj\omega + 1)$ about the 0-dg line.

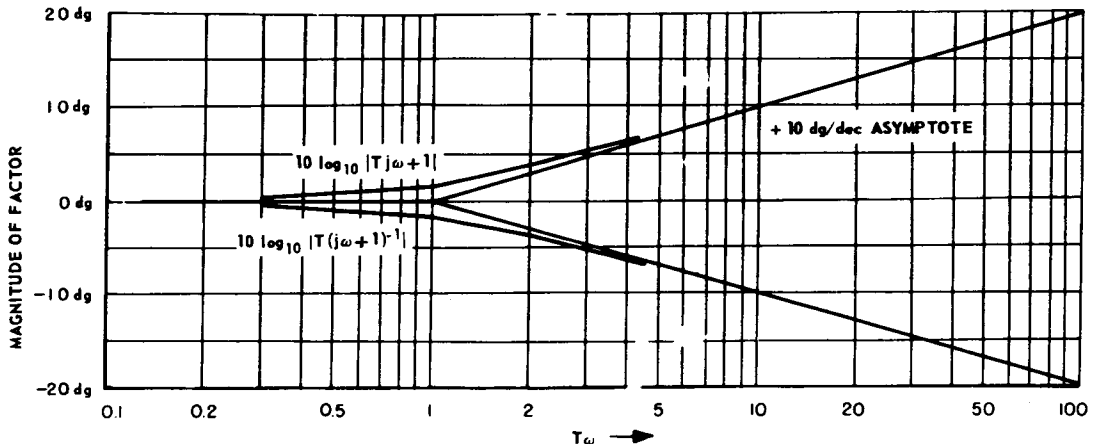


Fig. 5-7 Asymptotes and true magnitude curves for the first-order factor $(Tj\omega + 1)^{\pm 1}$.

THEORY

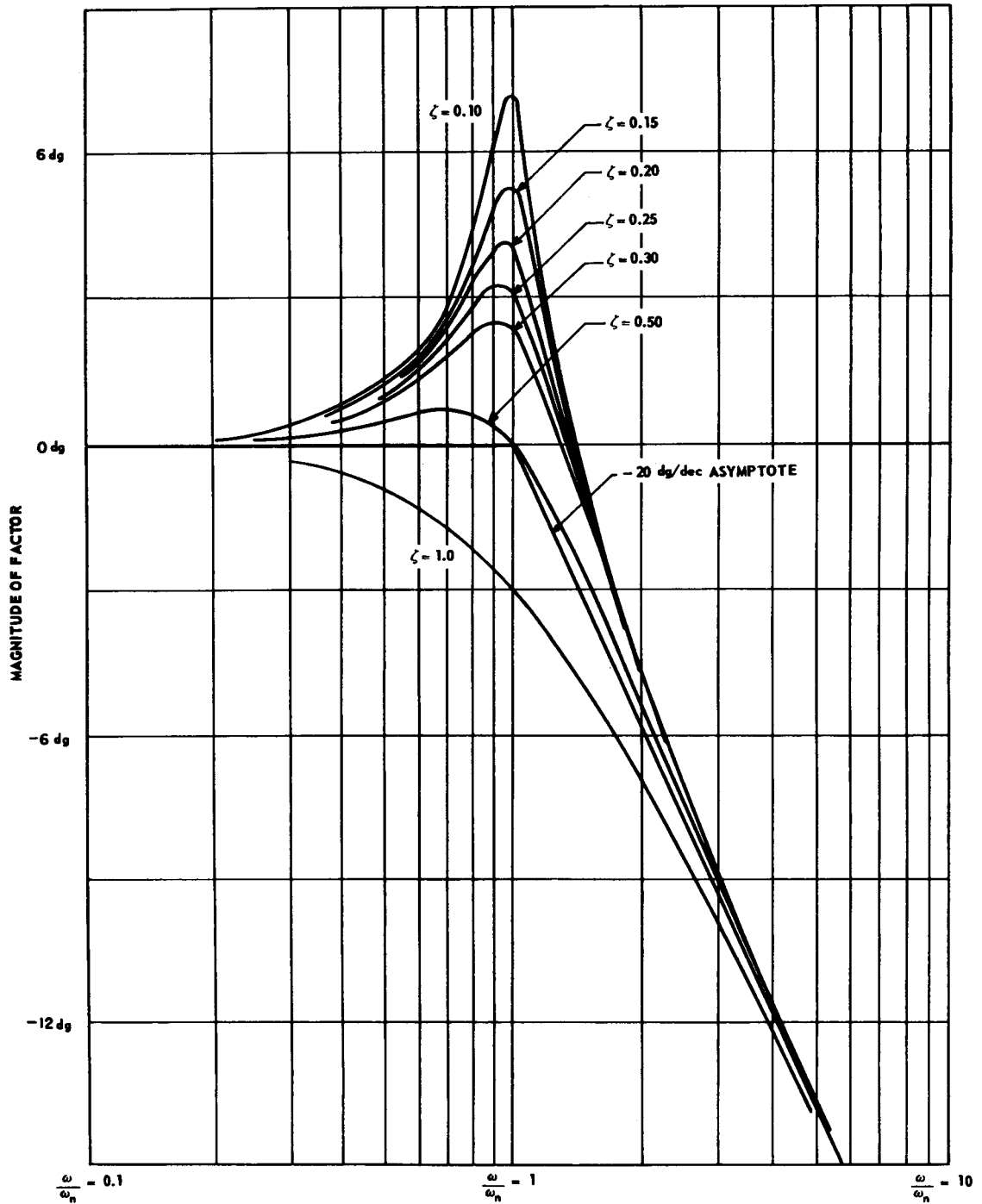


Fig. 5-8 Asymptotes and true curves for the second-order factor $\left[\left(i \frac{\omega}{\omega_n} \right)^2 + 2 \zeta_i \frac{\omega}{\omega_n} + 1 \right]^{-1}$.

GAIN DETERMINATION

The magnitude of the second-order factor

$$\left[\left(j \frac{\omega}{\omega_n} \right)^2 + 2\zeta j \frac{\omega}{\omega_n} + 1 \right]^{\pm 1}$$

can be approximated by two straight-line asymptotes. For $\omega \ll \omega_n$, the asymptote is the 0-dg line. For $\omega \gg \omega_n$, the asymptote is a straight line with a slope of ± 20 dg/decade crossing the 0-dg line at the break frequency, $\omega_b = \omega_n$. A set of second-order magnitude curves is shown in Fig. 5-8 for different values of the damping ratio ζ . Note that the approximation is best for $\zeta = 0.5$.

5-3.4 PHASE-ANGLE CURVES

The phase angle of the factor $(j\omega)^{\pm n}$ is a constant equal to $\pm 90n^\circ$. The phase-angle curves of the first-order factor $(Tj\omega + 1)^{\pm 1}$ are shown in Fig. 5-9. Note that each curve is symmetrical about the point on the curve

at which $\omega = 1/T$. The phase-angle curves of the second-order factor

$$\left[\left(j \frac{\omega}{\omega_n} \right)^2 + 2\zeta j \frac{\omega}{\omega_n} + 1 \right]$$

are shown in Fig. 5-10 for different values of the damping ratio ζ .

To plot the separate magnitude and phase-angle curves for a factored transfer function, separate plots are first made of the magnitude and phase-angle curves of each factor. In doing this, care must be taken to distinguish between numerator and denominator factors. Next, all the magnitudes are added at each frequency to obtain the composite magnitude curve. Similarly, all the phase angles are added at each frequency to obtain the composite phase-angle curve. It is important to note that the factors must be in the standard forms given in Eqs. (5-8), (5-9),

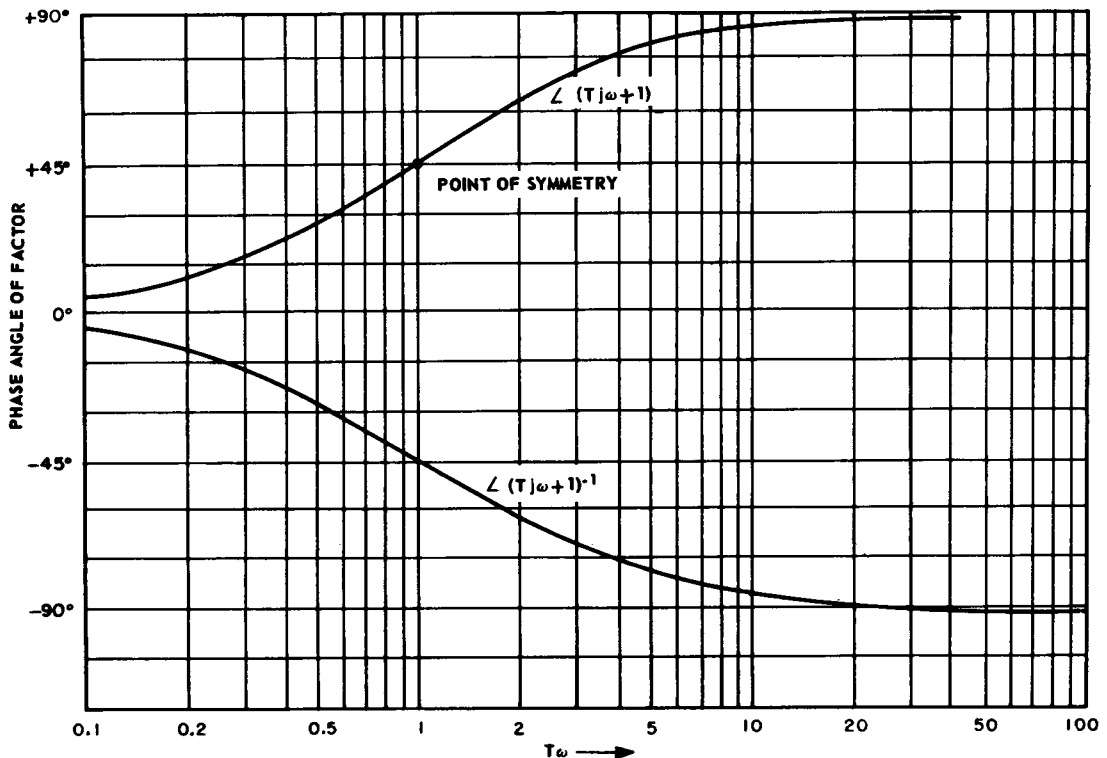


Fig. 5-9 Phase-angle curves for the first-order factor $(Tj\omega + 1)^{\pm 1}$.

THEORY

and (5-10) if the curves of Figs. 5-7 to 5-10 are to be used. The effect of the gain K can be incorporated by merely adding $10 \log_{10} K$ to the magnitude scale of the composite magnitude curve.

Example. The separate factors and composite curves for the function

$$G(j\omega) = K \frac{(0.2j\omega + 1)}{(j\omega) \left[\left(j \frac{\omega}{10} \right)^2 + 0.6 j \frac{\omega}{10} + 1 \right]} \quad (5-16)$$

where $K = 6.5$ are plotted in Figs. 5-11 and 5-12.

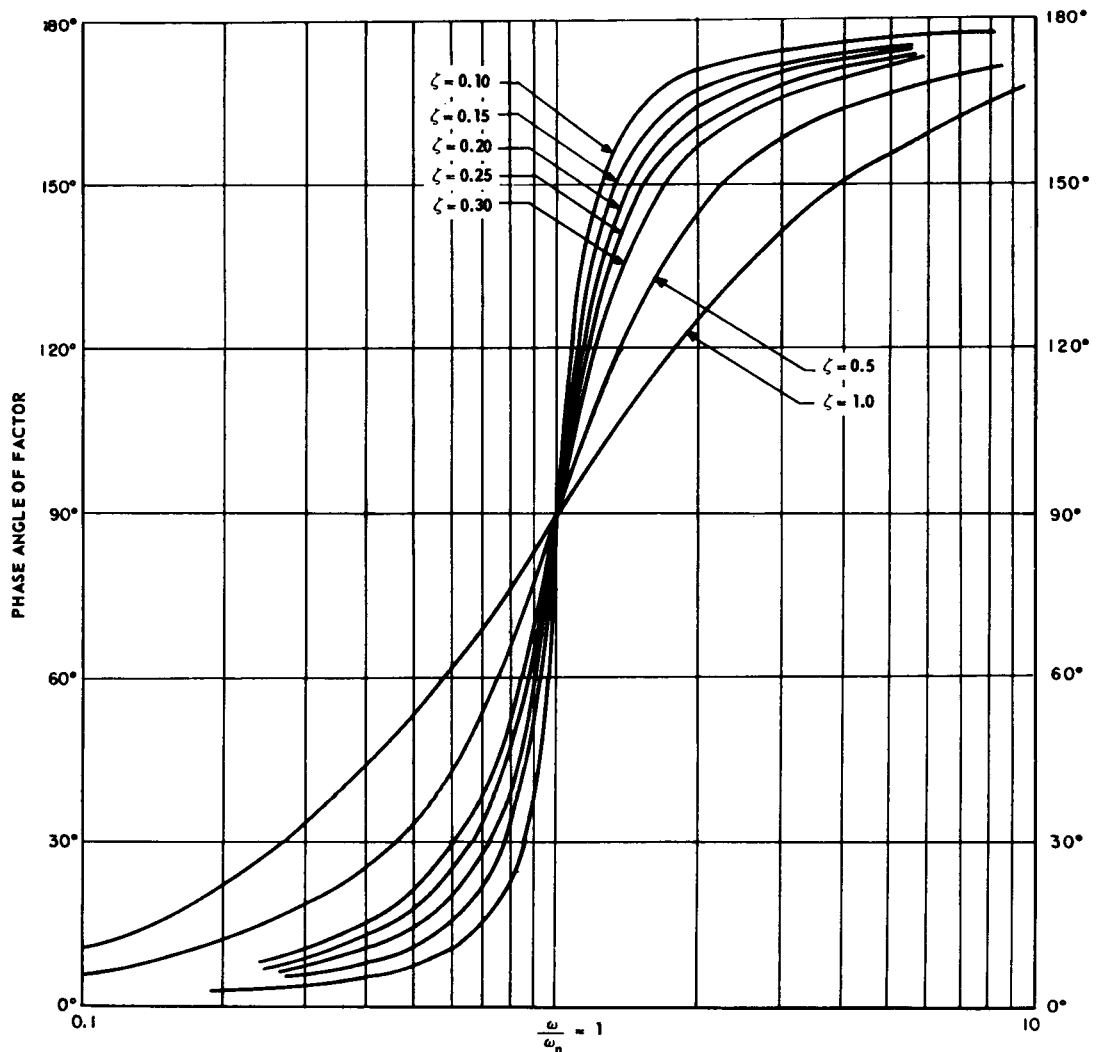


Fig. 5-10 Phase-angle curves for the second-order factor $\left[\left(j \frac{\omega}{\omega_n} \right)^2 + 2 \zeta j \frac{\omega}{\omega_n} + 1 \right]$.

GAIN DETERMINATION

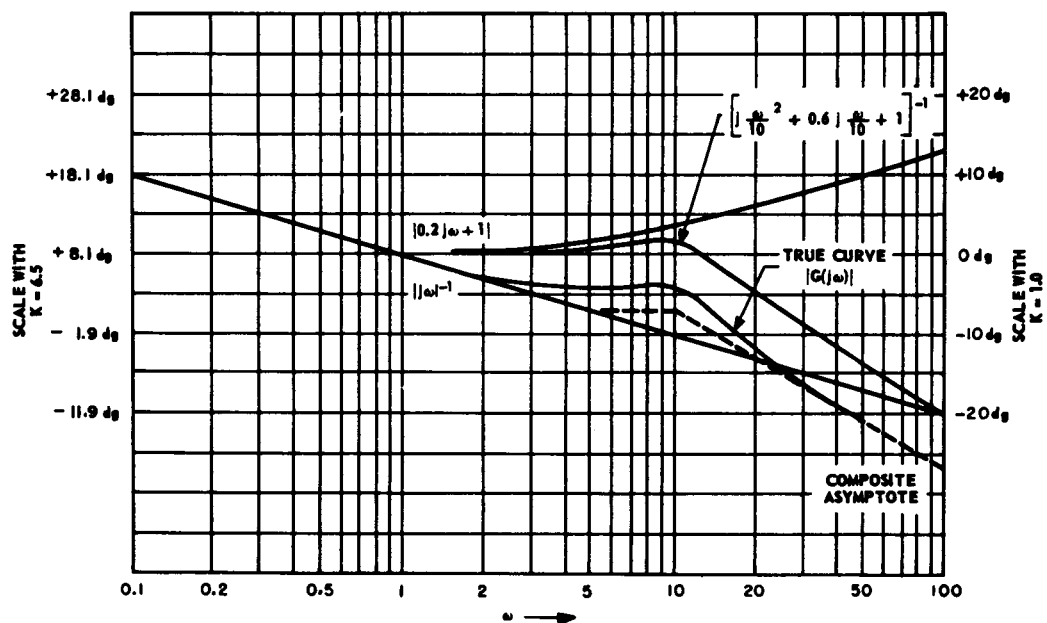


Fig. 5-11 Magnitude plots for $G(j\omega) = K \frac{[0.2j\omega + 1]}{j\omega \left[\left(j \frac{\omega}{10} \right)^2 + 0.6j \frac{\omega}{10} + 1 \right]}$.

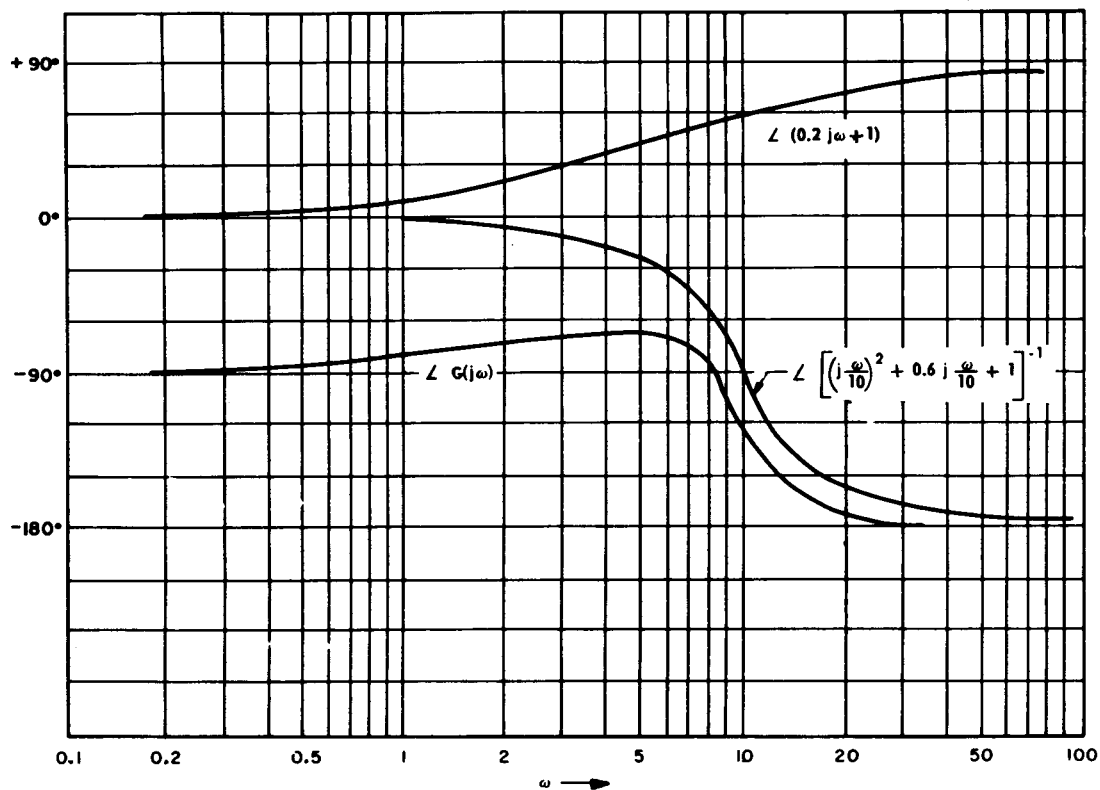


Fig. 5-12 Angle plots for $G(j\omega) = K \frac{(0.2j\omega + 1)}{j\omega \left[\left(j \frac{\omega}{10} \right)^2 + 0.6j \frac{\omega}{10} + 1 \right]}$.

THEORY

5-3.5 GAIN-PHASE PLANE

To facilitate design, a third method for representing frequency functions may be used. In this method, the magnitude and phase angle of a frequency function are plotted on a coordinate system called the *gain-phase plane*. The magnitude is plotted to a logarithmic scale (in decibels) and the phase angle is plotted to a linear scale. Frequency is the parameter for the gain-phase plot. The gain-phase plot can be determined directly from the frequency function by calculating the magnitude (in db) and phase angle of the function at various frequencies. Alternatively, the gain-phase plot can be determined through the intermediate use of the separate magnitude and phase-angle plots when the function to be plotted is in factored form. The gain-phase plot is most useful for determining the closed-loop response of a system from the open-loop response (see Pars. 5-4 and 5-5).

Example. The function

$$G(j\omega) = 6.5 \frac{(0.2j\omega + 1)}{(j\omega) \left[\left(j \frac{\omega}{10} \right)^2 + 0.6 j \frac{\omega}{10} + 1 \right]} \quad (5-17)$$

is plotted on the gain-phase plane in Fig. 5-13.

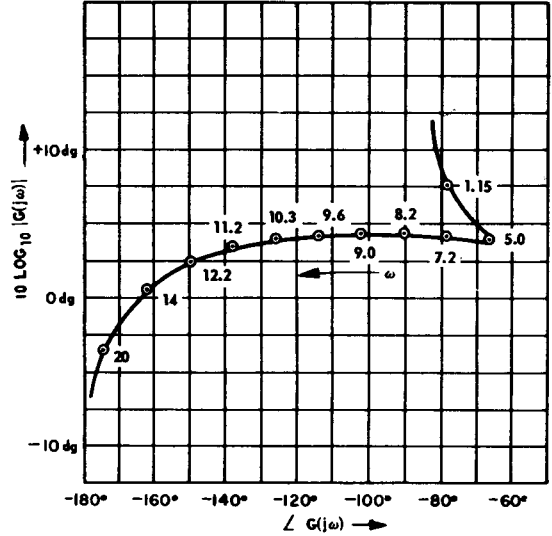


Fig. 5-13 Gain-phase plot of $G(j\omega) =$

$$6.5 \frac{(0.2j\omega + 1)}{j\omega \left[\left(j \frac{\omega}{10} \right)^2 + 0.6 j \frac{\omega}{10} + 1 \right]}$$

5-4 CLOSED-LOOP RESPONSE DETERMINATION^(4,5,9,12,13,14)

5-4.1 GENERAL

The relations that exist between the closed- and open-loop responses of a unity-feedback system can be obtained by considering the diagram in Fig. 5-1. In this diagram, it is clear that the open-loop responses is given by

$$\frac{C(j\omega)}{E(j\omega)} = G(j\omega) \quad (5-18)$$

The closed-loop response function $W(j\omega)$ is defined as follows:

$$\frac{C(j\omega)}{R(j\omega)} \triangleq W(j\omega) \quad (5-19)$$

Now, $E(j\omega) = R(j\omega) - C(j\omega)$. Substituting this expression for $E(j\omega)$ into Eq. (5-18), rearranging terms, and using the definition in Eq. (5-19), it is found that

$$W(j\omega) = \frac{G(j\omega)}{1 + G(j\omega)} \quad (5-20)$$

The transformation from the G plane to the W plane defined by Eq. (5-20) is used to determine the closed-loop response $W(j\omega)$ from the corresponding open-loop response $G(j\omega)$. Although a direct calculation of the W function is possible, this is often avoided

GAIN DETERMINATION

because it usually proves to be tedious. Instead, various aids for performing the G -to- W plane transformation are used. These are presented next.

5-4.2 POLAR-PLANE TECHNIQUE

In the polar-plane technique, a "vector" construction on the G plane is used to determine both the magnitude and phase angle of W . As an illustration, consider the G function sketched in Fig. 5-14. In this figure, once the $-1 + j0$ point is located, the following "vector" relations hold:

$$\overline{OB} = G \quad (5-21)$$

$$\overline{OA} = -1 \quad (5-22)$$

$$\overline{OB} - \overline{OA} = \overline{AB} = 1 + G \quad (5-23)$$

Then, the closed-loop response at each frequency can be determined from the construction of Fig. 5-14 as follows:†

$$|W(j\omega)| = \frac{|\overline{OB}|}{|\overline{AB}|} \quad (5-24)$$

$$\angle W(j\omega) = \angle ABO = \phi \quad (5-25)$$

The inverse G -plane construction for the closed-loop response is shown in Fig. 5-15. In this figure, the following "vector" relations hold:

$$\overline{OB} = G^{-1} \quad (5-26)$$

$$\overline{OA} = -1 \quad (5-27)$$

$$\overline{AO} + \overline{OB} = \overline{AB} = 1 + G^{-1} \quad (5-28)$$

The closed-loop response can be determined from the following relations:

$$|W(j\omega)| = \frac{|\overline{AO}|}{|\overline{AB}|} \quad (5-29)$$

$$\angle W(j\omega) = \angle OAB = \phi \quad (5-30)$$

To avoid "vector" constructions, constant magnitude and constant phase-angle contours that correspond to the G -to- W transformation of Eq. (5-20) are often used. The following definitions apply:

$$M = |W(j\omega)| \quad (5-31)$$

$$\phi = \angle W(j\omega) \quad (5-32)$$

$$N = \tan \phi \quad (5-33)$$

The transformation given in Eq. (5-20) can be used to map contours of constant M and constant N onto the G plane. The M contours appear as a set of bipolar circles as

†Symbol \angle denotes "angle"

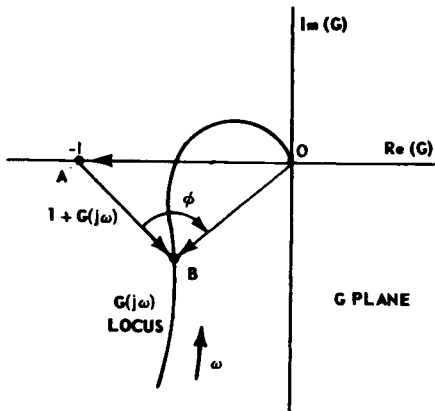


Fig. 5-14 Closed-loop response construction on the G plane.

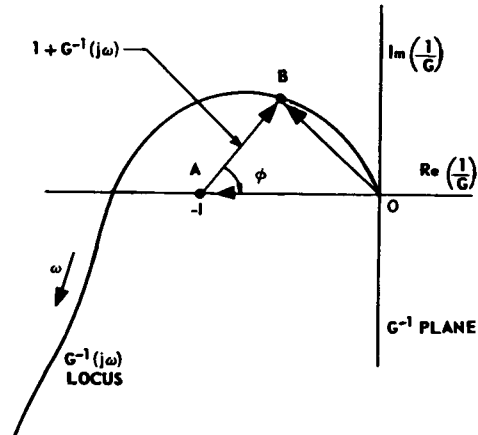


Fig. 5-15 Closed-loop response construction on the G^{-1} plane.

THEORY

shown in Fig. 5-16. The contours of constant N or ϕ are shown in Fig. 5-17.

To represent the constant M and ϕ contours in the G^{-1} plane, Eq. (5-20) is used. The M contours are a set of concentric circles and the ϕ contours are a set of straight lines. The M and ϕ contours for the G^{-1} plane are shown in Fig. 5-18.

The properties of the M and ϕ contours are listed in Table 5-1.

The M and ϕ contours are the lines or curves of constant M and constant ϕ as they appear in the G or G^{-1} planes. By constructing a chart of M and ϕ circles for the G or G^{-1} planes, one has the coordinate system of the W plane represented by circles and lines in the G or G^{-1} plane. The closed-loop magnitude M and

phase angle ϕ can be obtained directly from the G function by constructing the G function (or the G^{-1} function) on a chart of constant M and ϕ contours. At each frequency, the value of the M contour that intersects the G (or G^{-1}) function is the value of the magnitude of the closed-loop response W . Similarly, the value of the ϕ contour that intersects the G (or G^{-1}) function at a given frequency is the value of the phase angle of the closed-loop response W .

The M and ϕ contours aid greatly in performing the transformation from open- to closed-loop frequency response and are used to facilitate the design of a system when the shape of the G function is to be altered so as to improve performance.

TABLE 5-1 PROPERTIES OF M AND ϕ CONTOURS

| G Plane | G^{-1} Plane |
|--|--|
| <p style="text-align: center;">M contours</p> $y = Im(G)$ $x = Re(G)$ $y^2 + \left(x + \frac{M^2}{M^2 - 1}\right)^2 = \frac{M^2}{(M^2 - 1)^2}$ center: $\left(-\frac{M^2}{M^2 - 1}, 0\right)$ radius: $\left \frac{M}{M^2 - 1}\right $ intercept nearest origin: $-\frac{M}{M + 1}$ | <p style="text-align: center;">M contours</p> $y = Im(G^{-1})$ $x = Re(G^{-1})$ $y^2 + (1 + x)^2 = \frac{1}{M^2}$ center: $(-1, 0)$ radius: $\frac{1}{M}$ intercept nearest origin: $\frac{1 - M}{M}$ |
| <p style="text-align: center;">N contours</p> $N = \tan \phi$ $(x + 0.5)^2 + \left(y - \frac{1}{2N}\right)^2 = \frac{1}{4} \left(\frac{N^2 + 1}{N^2}\right)$ center: $\left(-\frac{1}{2}, \frac{1}{2N}\right)$ radius: $\frac{1}{2} \frac{\sqrt{N^2 + 1}}{N}$ | <p style="text-align: center;">N contours</p> $N = \tan \phi$ $y + Nx + N = 0$ <p>Note: N contours are a family of radial lines emanating from the center of the M circles</p> |

GAIN DETERMINATION

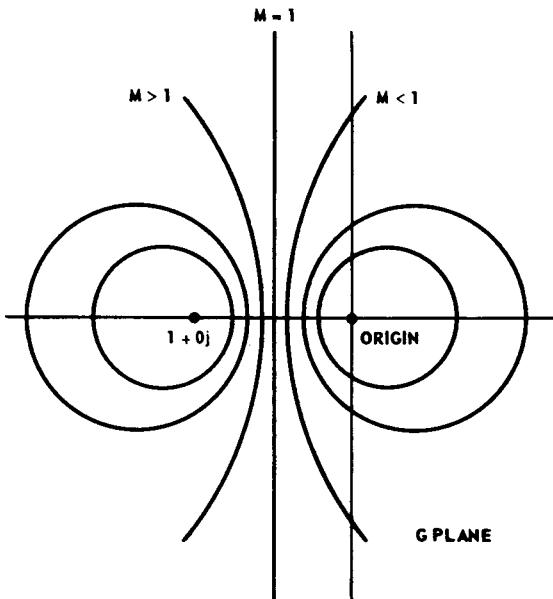


Fig. 5-16 Contours of constant M in the G plane.

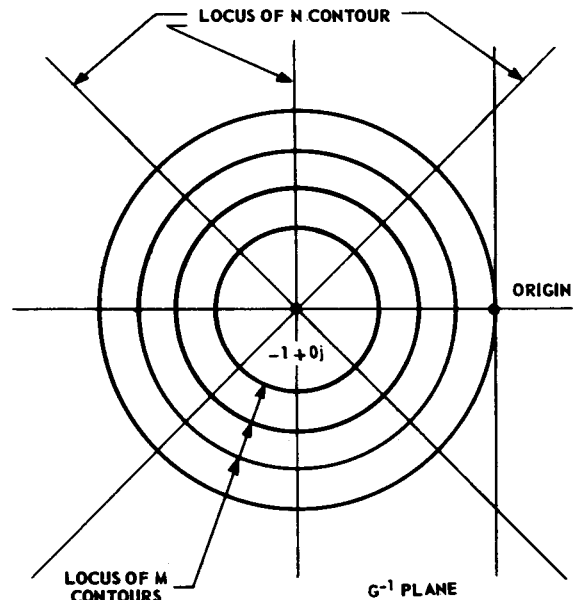


Fig. 5-18 Contours of constant M and constant ϕ in the G^{-1} plane.

5-4.3 GAIN-PHASE PLANE TECHNIQUE (NICHOLS CHART)

Since constructions on the gain-phase plane involving cascaded functions and gain alterations are usually simpler than similar constructions on the polar plane, a chart of constant M and ϕ contours has been constructed for the gain-phase plane. This chart is called the Nichols chart and is shown in Figs. 5-19 and 5-20. Figure 5-19 presents a

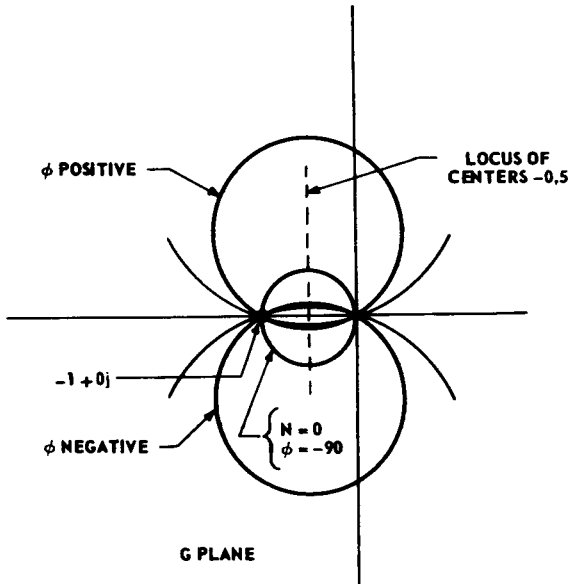


Fig. 5-17 Contours of constant phase in the G plane.

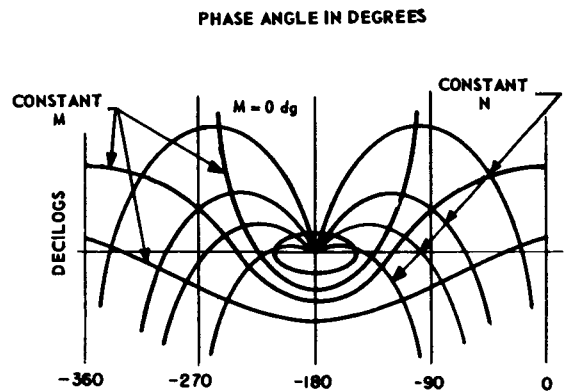


Fig. 5-19 Chart showing symmetry of M - N contours about phase of 180 degrees (Nichols Chart).

Reprinted with permission from *Principles of Servomechanisms*, by D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

THEORY

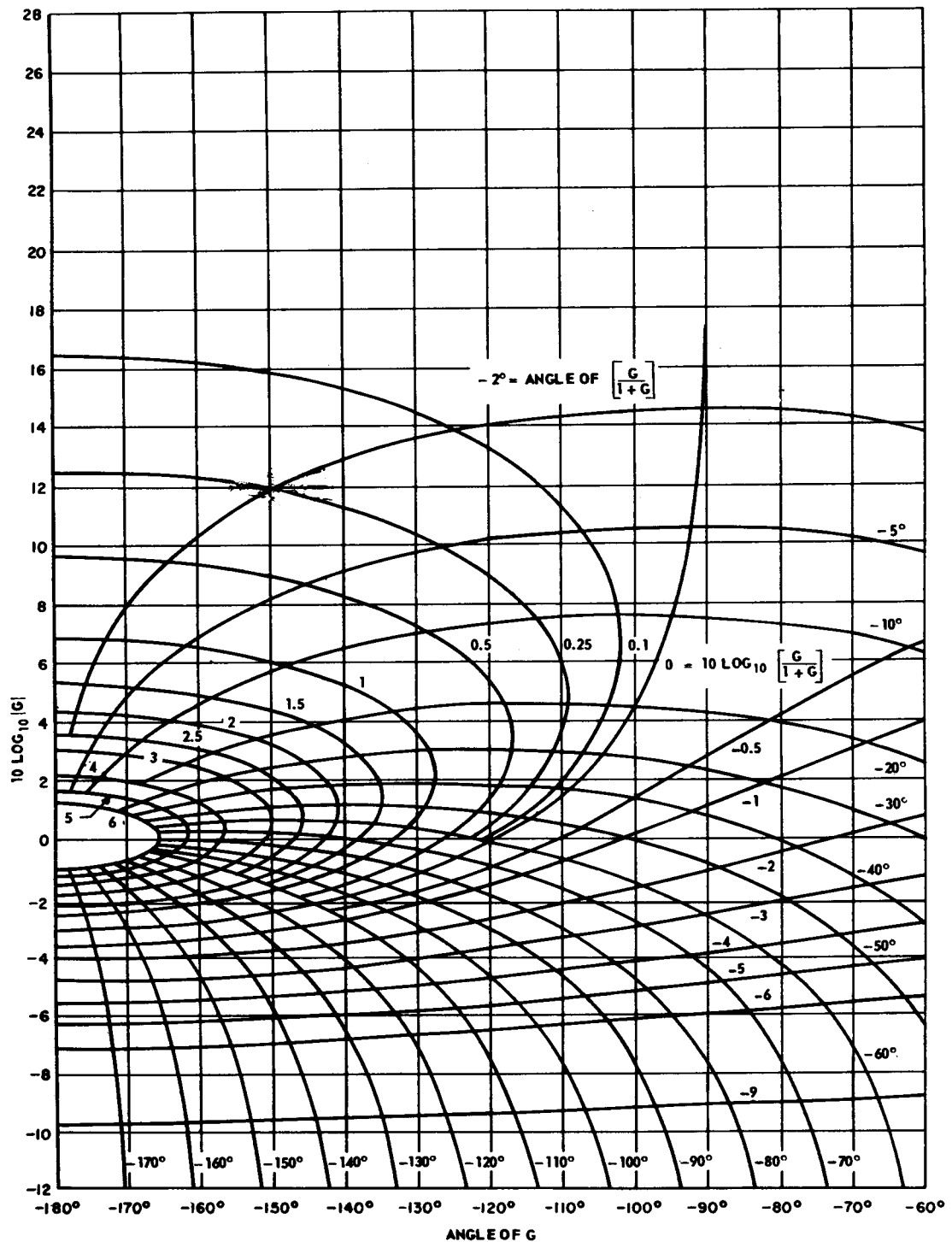


Fig. 5-20 Nichols chart.

GAIN DETERMINATION

large-scale view of the *Nichols* chart and Fig. 5-20 presents only that part of the Nichols chart that is most useful for design purposes.

The Nichols chart is used in the same way that the *M-N* contours are used on the polar plane. The G function is plotted on the Nichols chart. The value of the M contour that intersects the G function at a given frequency is the value of the magnitude of the closed-loop response W at that frequency. Similarly, the intersection of the G function with the ϕ contours determines the phase angle of the closed-loop response W as a function of frequency.

5-4.4 NONUNITY-FEEDBACK SYSTEMS

If the closed-loop response of a nonunity-feedback system is sought, a slight modification of the procedure used for the unity-feedback system will enable the designer to use the Nichols chart and the polar *M-N* contours as well.

The closed-loop response of the nonunity-feedback system (Fig. 5-21) can be written as follows :

$$\frac{C(j\omega)}{R(j\omega)} = \frac{1}{H(j\omega)} \left[\frac{G(j\omega) H(j\omega)}{1 + G(j\omega) H(j\omega)} \right] \quad (5-34)$$

Since the bracketed portion of the right-hand side of Eq. (5-34) has the same form as the right-hand side of Eq. (5-20), the Nichols chart (or the polar *M-N* contours) can be used to find $GH/(1 + GH)$ from a plot of $G(j\omega) H(j\omega)$. The closed-loop response $C(j\omega)/R(j\omega)$ can then be found by multiplying $GH/(1 + GH)$ by H^{-1} at each frequency.

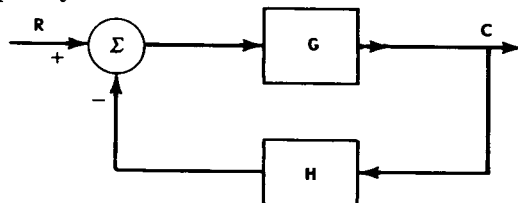


Fig. 5-21 Nonunity-feedback system.

5-5 SETTING THE GAIN FOR A SPECIFIED M_p ^(4,5,9,12,13,14)

5-5.1 GENERAL

A primary problem encountered in servo system design is the determination of the loop gain K required to produce a specified degree of stability. For a unity-feedback system (Fig. 5-1), the stability of the system is determined by the location of the $G(j\omega)$ locus with respect to the point $-1 + j0$ (see Nyquist criterion, Par. 4-3). For a nonunity-feedback system (Fig. 5-21), however, the stability of the system is determined by the location of the $G(j\omega) H(j\omega)$ locus with respect to the point $-1 + j0$. One analytical approach can serve for both types of systems if it is noted that, by redrawing Fig. 5-21, the study of the stability of a nonunity-feedback system can be expressed in terms

of the stability of a unity-feedback system cascaded with another transfer function (Fig. 5-22). Thus, the discussion of stability can be limited to unity-feedback systems.

A system is said to have a low degree of stability if the normal mode of response is highly oscillatory. Such a system is also said

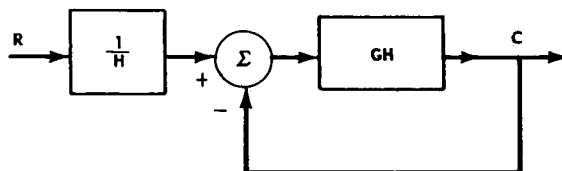


Fig. 5-22 System equivalent to the system of Fig. 5-21.

THEORY

to have a low relative stability. The degree of stability of a stable unity-feedback system can be measured by the closeness of approach of the $G(j\omega)$ locus to the point $-1 + j0$. An examination of the polar M contours (Figs. 5-16 and 5-18) or the Nichols chart (Fig. 5-20) shows that, the larger the value of M , the more closely the $G(j\omega)$ contour approaches the point $-1 + j0$. The peak magnitude of the closed-loop response $W(j\omega)$ is called M_p . By limiting this peak value, the degree of stability of a system can be maintained within a specified bound.

If $G(j\omega)$ is specified, except for a factor K , the degree of stability of the closed-loop response $W(j\omega)$ corresponding to this $G(j\omega)$ can be changed by adjusting the value of K . If the degree of stability as measured by M_p is specified, K is uniquely determined. The determination of K for a specified M_p is usually accomplished by a graphical construction in the polar or gain-phase plane.

5-5.2 POLAR-PLANE CONSTRUCTION

The polar-plane construction required to determine K [the gain of $G(j\omega)$] for a specified M_p is shown in Fig. 5-23 for the G/K plane and Fig. 5-24 for the KG^{-1} plane.

The procedure used for a construction on the direct, or G/K , plane (Fig. 5-23) is as follows:

(a) $\frac{G(j\omega)}{K}$ is plotted as a function of ω .

(b) A straight line is drawn from the origin, making an angle ψ with the real axis, where

$$\psi = \sin^{-1} \left(\frac{1}{M_p} \right)$$

This line is called the ψ line.

(c) A circle with center on the real axis is constructed, tangent to the ψ line and the $G(j\omega)/K$ locus.

(d) A line is drawn from the point of tangency of the ψ line with the circle (point b in Fig. 5-23) normal to the real axis.

(e) The value of $\text{Re}(G/K)$ at the point of intersection of the normal with the real axis (point a in Fig. 5-23) is the reciprocal of the gain K which must multiply $G(j\omega)/K$ to produce the specified M_p .

(f) The angular frequency at the point of tangency of the $G(j\omega)/K$ locus with the circle is the resonant frequency of the closed-loop system having the specified M_p (ω_R in Fig. 5-23).

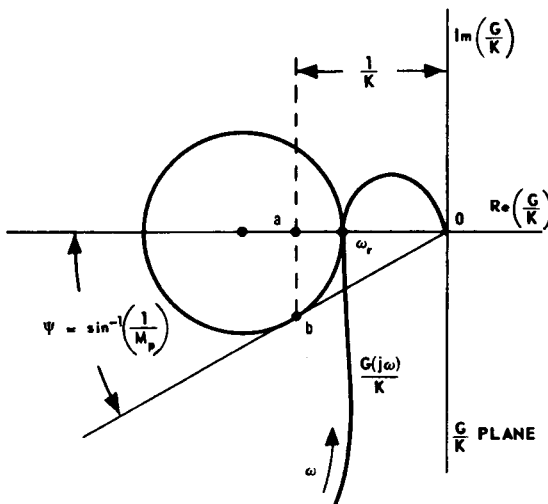


Fig. 5-23 Construction for gain determination on direct (G/K) plane.

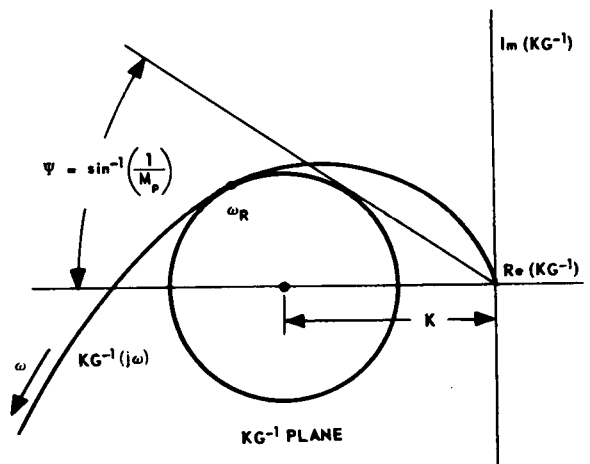


Fig. 5-24 Construction for gain determination on inverse (KG^{-1}) plane.

GAIN DETERMINATION

The procedure used for a construction on the inverse, or KG^{-1} , plane (Fig. 5-24) is as follows:

- (a) $KG^{-1}(j\omega)$ is plotted as a function of ω .
- (b) A straight line (ψ line) is drawn from the origin, forming an angle ψ with the real axis, where

$$\psi = \sin^{-1} \left(\frac{1}{M_p} \right)$$

- (c) A circle with center on the real axis is constructed, tangent to the KG^{-1} locus and the ψ line.

- (d) The center of the circle is the point $-K + j0$. Thus, the coordinate of the center of the circle on the real axis is the value of K used to multiply $G(j\omega)$ to produce the desired M_p .

- (e) The angular frequency at the point of tangency of the circle with the $KG^{-1}(j\omega)$

locus is the resonant frequency of the closed-loop system having the specified M_p (ω_R in Fig. 5-24).

Example. A unity-feedback system has the following open-loop frequency response:

$$G(j\omega) = \frac{K}{j\omega[(j\omega)^2 + 0.6j\omega + 1]} \quad (5-35)$$

Find K and ω_R for $M_p = 1.6$.

Solution.

- (a) Direct-plane procedure:

- (1) The $\frac{G(j\omega)}{K}$ locus is plotted (Fig. 5-25).

- (2) The ψ line is drawn with $\psi = \sin^{-1}$

$$\left(\frac{1}{1.6} \right) = 38.7^\circ.$$

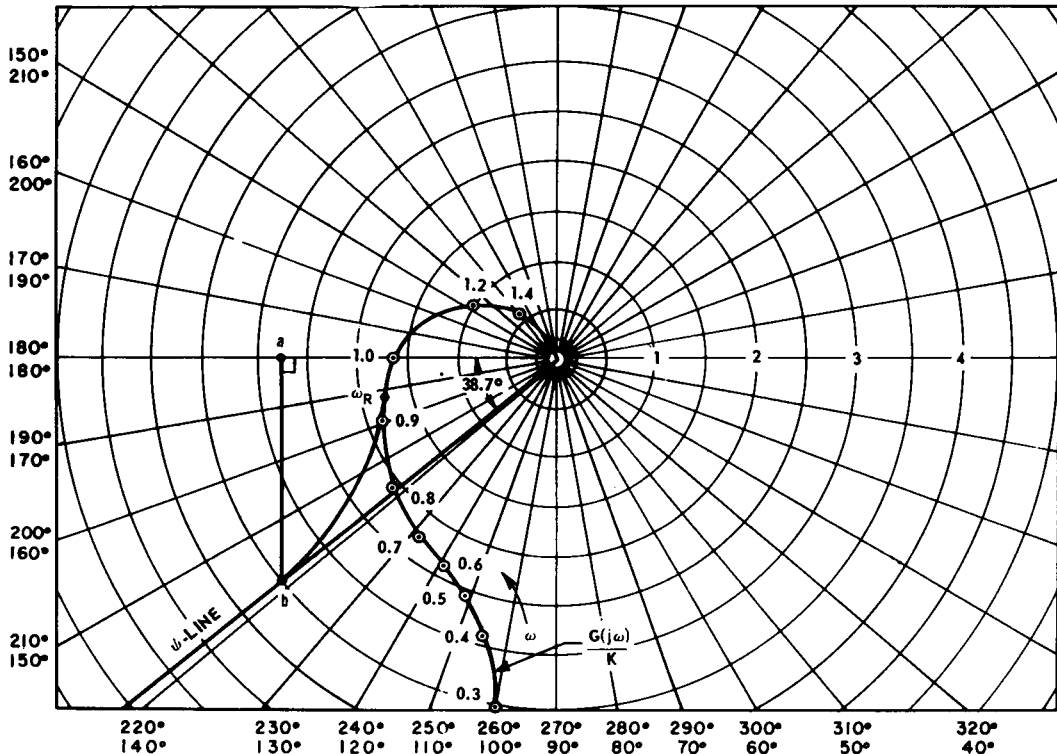


Fig. 5-25 Direct-plane determination of K for $M_p = 1.6$, $G(j\omega) = \frac{K}{j\omega[(j\omega)^2 + 0.6j\omega + 1]}$.

THEORY

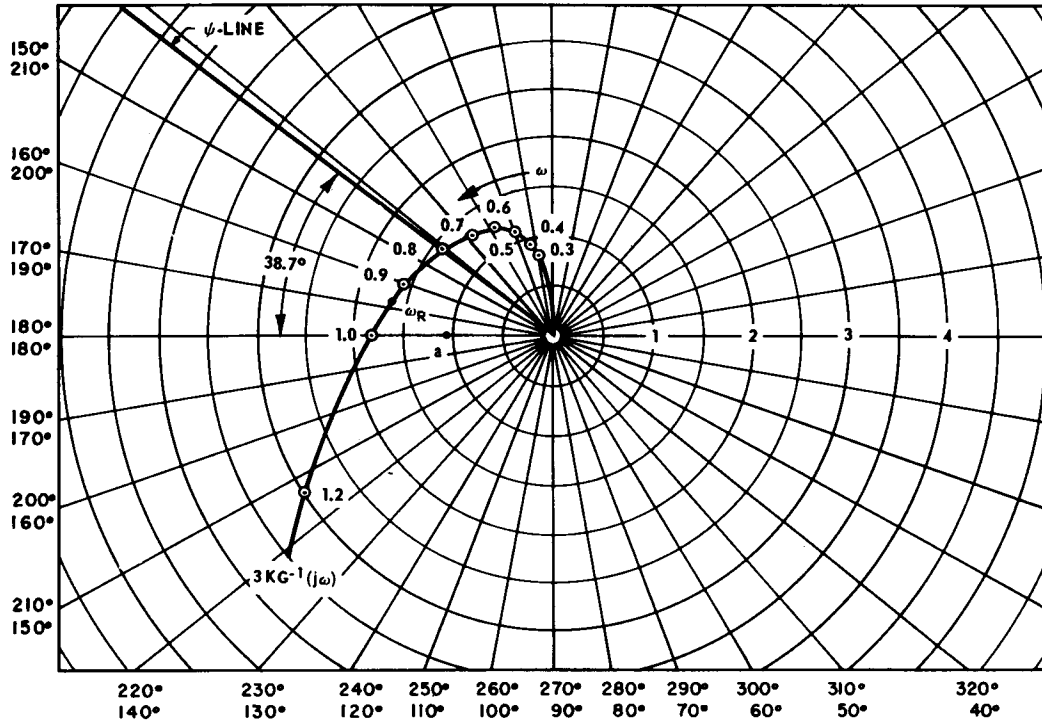


Fig. 5-26 Inverse-plane determination of K for $M_p = 1.6$, $G(j\omega) = \frac{K}{j\omega [(j\omega)^2 + 0.6j\omega + 1]}$

(3) The circle with center on the real axis, tangent to the $G(j\omega)/K$ locus and the ψ line, is constructed.

(4) A line is drawn from point b perpendicular to the real axis, intersecting the real axis at the point $-2.78 + j0$ (point a).

(5) Thus, $\frac{1}{K} = 2.78$, or $K = 0.36$.

(6) The resonant frequency is $\omega_R = 0.94$ rad/sec.

(b) Inverse-plane procedure:

(1) The locus of $3KG^{-1}(j\omega)$ is plotted (Fig. 5-26).

(2) The ψ line for $M_p = 1.6$ ($\psi = 38.7^\circ$) is drawn.

(3) The tangent circle is constructed.

(4) The center of the tangent circle is at $1.08 + j0$ (point a).

(5) Thus, $3K = 1.08$, or $K = 0.36$.

(6) The resonant frequency is $\omega_R = 0.94$ rad/sec.

5-5.3 GAIN-PHASE PLANE CONSTRUCTION

The construction required to determine the gain K for a specified M_p is simpler when the gain-phase plane rather than the polar plane is employed. In the gain-phase plane construction, changing the gain merely moves the $G(j\omega)$ locus in a vertical direction without changing the phase angle. The gain-phase plane construction is carried out as follows (Fig. 5-27):

(a) The $G(j\omega)/K$ locus is drawn on the gain-phase plane.

(b) The $G(j\omega)/K$ locus is placed over the Nichols chart or, more specifically, a plot of the desired M_p contour is made on the gain-phase plane. The two coordinate systems are then aligned so that angles coincide.

(c) The $G(j\omega)/K$ locus is moved up (or down) until it is tangent to the specified M_p contour.

GAIN DETERMINATION

(d) The intersection of the 0-dg line of the M_p contour with the $G(j\omega)/K$ magnitude scale gives the value of $10 \log_{10} K$, where K is the value of the gain by which $G(j\omega)/K$ must be multiplied to produce the specified M_p .

(e) The angular frequency at the point of tangency of the M_p contour with the $G(j\omega)/K$ locus determines the resonant frequency of the closed-loop system having the specified M_p .

Example. A unity-feedback system has the open-loop frequency response

$$G(j\omega) = \frac{K(0.2j\omega + 1)}{j\omega \left[\left(j\frac{\omega}{10} \right)^2 + 0.6j\frac{\omega}{10} + 1 \right]} \quad (5-36)$$

Find K and ω_R for $M_p = 1.5$.

Solution.

(a) The $\frac{G(j\omega)}{K}$ locus is plotted (Fig. 5-28).

(b) The $G(j\omega)/K$ locus is placed over the $M_p = 1.5$ contour, the phase-angle coordinates

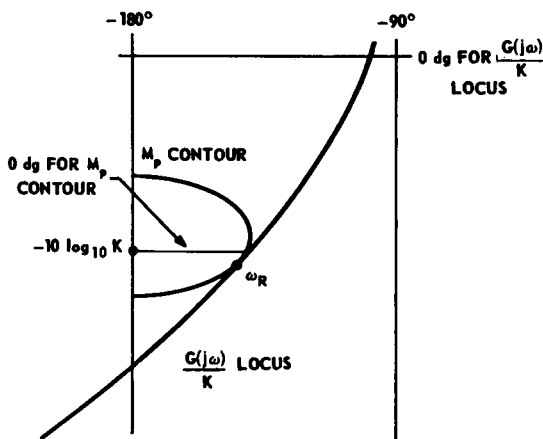


Fig. 5-27 Construction for gain determination on gain-phase plane.

are aligned, and the locus is moved vertically until tangency occurs.

(c) The point of tangency occurs at $\omega_R = 12 \text{ rad/sec}$.

(d) The intersection of the 0-dg line of the M_p contour with the magnitude scale of $G(j\omega)/K$ yields

$$-10 \log_{10} K = -4.15 \text{ dg, or } K = 2.6$$

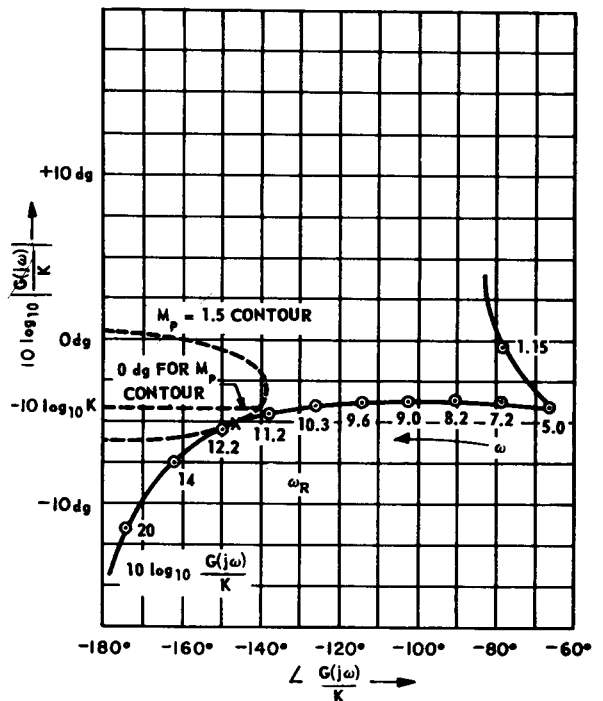


Fig. 5-28 Gain-phase plane determination of K for $M_p = 1.5$,

$$G(j\omega) = K \frac{(0.2j\omega + 1)}{j\omega \left[\left(j\frac{\omega}{10} \right)^2 + 0.6j\frac{\omega}{10} + 1 \right]}$$

5-6 APPROXIMATE PROCEDURES^(4,5,9)

5-6.1 PHASE MARGIN AND GAIN MARGIN

The peak magnitude of the closed-loop response is not the only measure of the degree of stability that is commonly used. More direct, but less reliable, descriptions of the approach of the $G(j\omega)$ locus to the point $-1+j0$ are available. These measures of the degree of stability are called *phase margin* and *gain margin*.

5-6.2 Phase Margin (Fig. 5-29)

The phase margin (p.m.) of the open-loop function $G(j\omega)$ of a unity-feedback system equals $[180^\circ + \angle G(j\omega)]$ at the frequency for which the magnitude of $G(j\omega)$ is unity.

5-6.3 Gain Margin (Fig. 5-29)

The gain margin (g.m.) of the open-loop function $G(j\omega)$ of a unity-feedback system is the reciprocal of the magnitude of $G(j\omega)$ at the frequency for which the angle of $G(j\omega)$ is -180° .

The primary advantage of the use of the phase margin or the gain margin as a measure of the degree of stability is that calculations may be made directly on the separate magnitude and phase-angle plots.

In practice, the phase margin is used more widely than the gain margin as a degree-of-stability criterion, while the gain margin that results when the phase margin is specified is used as a measure of the goodness of performance. A system with a low gain margin is considered to have a poor performance.

The usual ranges of phase margin and gain margin for which performance will probably be satisfactory are the following:

$$30^\circ < \text{p.m.} < 60^\circ \quad (5-37)$$

$$2.5 < \text{g.m.} < 10 \quad (5-38)$$

When the phase margin is used as a degree-of-stability criterion for setting the gain K of a unity-feedback system, the procedure is developed directly from the definition of phase margin as follows:

(a) The separate amplitude and phase plots (or the gain-phase plot) of $G(j\omega)/K$ are constructed.

(b) The frequency at which

$$\angle \frac{G(j\omega)}{K} = -180^\circ + \text{p.m.} \quad (5-39)$$

is determined.

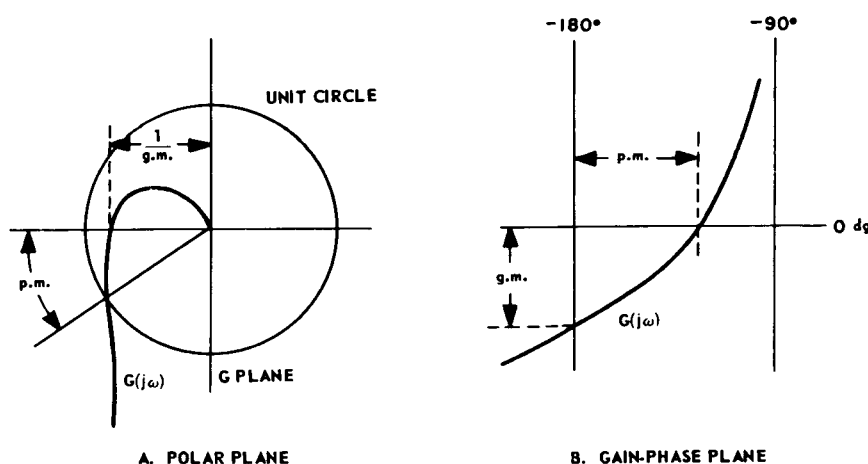


Fig. 5-29 Phase margin and gain margin.

GAIN DETERMINATION

(c) At this frequency, $|G(j\omega)/K|$ is determined.

(d) At this frequency, K is chosen such that

$$|G(j\omega)| = 1 \quad (5-40)$$

Note that, if only a rough approximation is desired, the asymptotic magnitude curve may be used rather than the true magnitude curve.

Example.

The function plotted in Figs. 5-12 and 5-13 is the open-loop function of a unity-feedback system. The phase margin of the system is to be set at 45° . In Fig. 5-13, $\omega = 11$ rad/sec when

$$\angle \frac{G(j\omega)}{K} = -180^\circ + 45^\circ = -135^\circ \quad (5-41)$$

In Fig. 5-11, for $\omega = 11$ rad/sec

$$10 \log_{10} \left| \frac{G(j\omega)}{K} \right| = -4.5 \text{ dg} \quad (5-42)$$

To have $|G(j\omega)| = 1$ at $\omega = 11$ rad/sec

$$10 \log_{10} K = 4.5 \text{ dg, or } K = 2.82 \quad (5-43)$$

Note, in this example, that the use of the asymptotic curve to estimate K gives a poor result. The magnitude of the asymptotic approximation for $G(j\omega)/K$ at $\omega = 11$ rad/sec is -7.5 dg. This would give an approximate value of $K = 5.61$ for a 45° phase margin. The error of approximation is a factor of two, which is too large to be acceptable. One should note further that, for this system, the gain margin is infinite since the negative phase shift never exceeds 180° .

5-6.4 GENERAL COMMENTS ON THE PHASE-MARGIN CRITERION

The phase-margin criterion used as a measure of the degree of stability is a good substitute for the M_p criterion provided that the G function does not have low damping-ratio quadratic factors ($\zeta \leq 0.3$) with natural frequencies in the range where

$$-135^\circ < \angle G(j\omega) < -225^\circ \quad (5-44)$$

If no low damping-ratio quadratics are present, then the gain determined from the true magnitude curve (or the asymptotic

magnitude curve) for a specified phase margin is a good approximation to the gain determined from the corresponding M_p criterion. The M_p criterion corresponding to a given phase margin may be found from the relation

$$\text{p.m.} \approx \sin^{-1} \left(\frac{1}{M_p} \right) \quad (5-45)$$

The frequency at which the phase margin is determined is: (1) the magnitude cross-over frequency ω_{cm} if the true magnitude curve is used; and (2) the asymptote cross-over frequency ω_c if the asymptotic magnitude curve is used. For the M_p corresponding to the specified phase margin, the frequencies ω_{cm} or ω_c are good approximations to the resonant frequency of the system ω_R .

5-6.5 APPROXIMATE CLOSED-LOOP RESPONSE

If the phase-margin criterion is used in conjunction with the separate amplitude and phase-angle plots, a rapid estimation of the closed-loop magnitude response is obtainable by means of the following relations:

(a) For a unity-feedback system (Fig. 5-1)

$$|W(j\omega)| \approx 1, \quad \text{when } |G(j\omega)| \gg 1 \quad (5-46)$$

$$|W(j\omega)| \approx |G(j\omega)|, \quad \text{when } |G(j\omega)| \ll 1 \quad (5-47)$$

(b) For a nonunity-feedback system (Fig. 5-21)

$$\left| \frac{C(j\omega)}{R(j\omega)} \right| \approx \left| \frac{1}{H(j\omega)} \right|, \quad \text{when } |G(j\omega) H(j\omega)| \gg 1 \quad (5-48)$$

$$\left| \frac{C(j\omega)}{R(j\omega)} \right| \approx |G(j\omega)|, \quad \text{when } |G(j\omega) H(j\omega)| \ll 1 \quad (5-49)$$

In the approximate equations (5-46) to (5-49), the boundary is always the point where the magnitude of the open-loop function is unity. Since this point is determined

THEORY

directly in the phase-margin procedure, the approximate closed-loop response for a given phase-margin criterion can be constructed as follows:

(a) For unity feedback, the magnitude crossover frequency ω_{cm} is determined by means of the phase-margin criterion.

(b) Usually for $\omega < \omega_{cm}$, $|G(j\omega)| > 1$, and for $\omega > \omega_{cm}$, $|G(j\omega)| < 1$ [$G(j\omega)$ is monotonic].

(c) Therefore, for $\omega < \omega_{cm}$, $|W(j\omega)| \approx 1$. For $\omega > \omega_{cm}$, $|W(j\omega)| \approx |G(j\omega)|$.

(d) At $\omega = \omega_{cm}$, $|W(j\omega)| \approx M_p$, where M_p is determined for the specified phase margin from Eq. (5-45).

(e) From the high-frequency ($\omega > \omega_{cm}$) and low-frequency ($\omega < \omega_{cm}$) behavior together with the behavior at $\omega = \omega_{cm}$, the entire magnitude response $|W(j\omega)|$ may be approximated.

If $|G(j\omega)|$ is not monotonic as defined in step (b), several magnitude crossover points will exist and Eqs. (5-46) and (5-47) must be used directly. In this case, the approximation should not be trusted unless the crossover points are widely separated (at least 1 decade apart) or $|G(j\omega)| \gg 1$ or $\ll 1$ between the crossover points.

The procedure for nonunity-feedback systems is similar to that described here for unity-feedback systems and is based on Eqs. (5-48) and (5-49).

If only a very rough approximation is desired, the asymptotic magnitude curves may always be used to reduce calculation time.

Example. The open-loop function of a unity-feedback system is

$$G(j\omega) = \frac{K}{j\omega(j\omega + 1)} \quad (5-50)$$

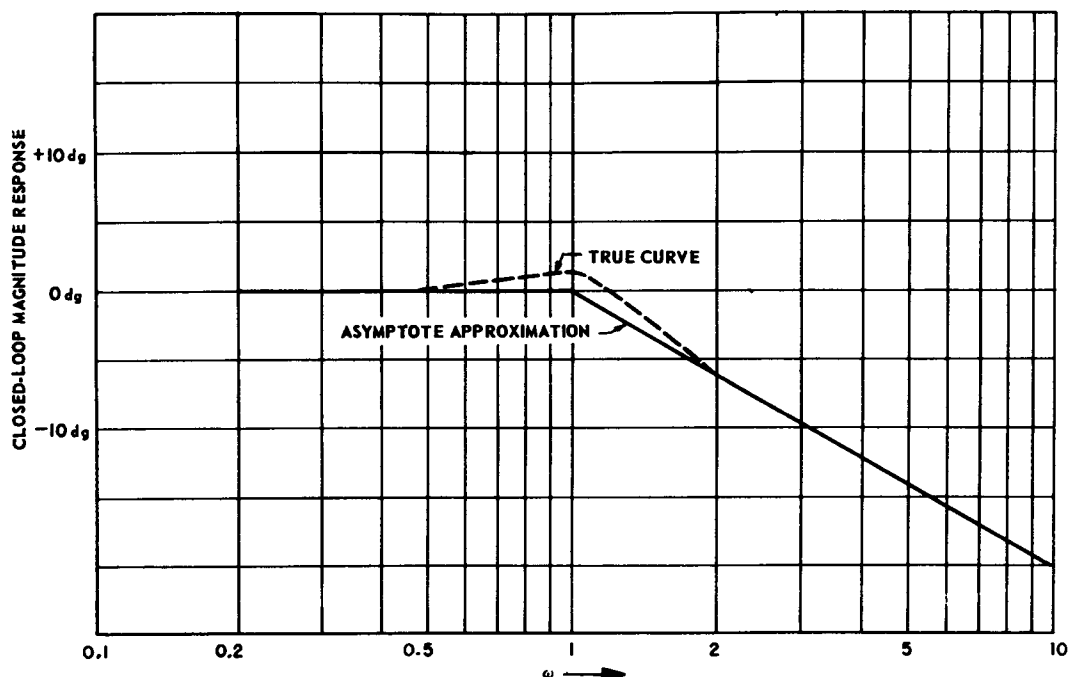


Fig. 5-30 Approximate closed-loop magnitude response of unity-feedback system,

$$G(j\omega) = \frac{K}{j\omega(j\omega + 1)}.$$

GAIN DETERMINATION

The gain K is to be set for a phase margin of 45° . Find K , ω_{cm} , and the approximate closed-loop magnitude response $W(j\omega)$.

Solution.

$$(a) \angle \frac{G(j\omega)}{K} = -180^\circ + 45^\circ = -135^\circ$$

for $\omega = 1$ (the crossover frequency).

$$(b) \text{ At } \omega = 1, \frac{G(j\omega)}{K} = 0.707.$$

$$(c) \text{ Therefore, } K = 1.41 \text{ and } \omega_{cm} = 1.$$

(d) If the asymptotic magnitude curve for $G(j\omega)/K$ is used, the asymptotic magnitude

of $G(j\omega)/K = 1$ at $\omega = 1$. Thus, use of the asymptotic magnitude rather than the true magnitude produces an error of 40 percent in the determination of K .

(e) If the approximation to the closed-loop magnitude response is based on the asymptotic magnitude curve of $G(j\omega)/K$, then for $\omega < \omega_{cm}$, the magnitude of $W(j\omega)$ is unity. For $\omega > \omega_{cm}$, the magnitude of $W(j\omega)$ is represented by a straight line with a slope of -20 dg/decade, crossing 0 dg at $\omega = \omega_{cm} = 1$. The M_p corresponding to a phase margin of 45° is 1.41, or 1.5 dg. The approximate closed-loop magnitude response is shown in Fig. 5-30.

5-7 ROOT-LOCUS METHOD^(7,8)

5-7.1 GENERAL

The root-locus method deals primarily with the study of the motion in the s plane of the roots of the characteristic equation of a system as a function of the gain K . The relation between stability and gain can be observed directly through use of this method by noting how the roots move from the left half of the s plane (stable roots) to the right half of the s plane (unstable roots) as K is varied. A system can be characterized as having a low degree of stability if its roots lie in the left half of the s plane but are very close to the imaginary axis.

Gain determination by means of the root-locus method is based on the fact that many practical systems have a pair of complex closed-loop poles that are closer to the origin than any other complex poles of the system. These poles are called the *dominant* poles of the system. By assigning a specified value to a characteristic parameter of the dominant pole pair, the gain of the system may be fixed by a measure of the degree of stability related to the dominant pole pair.

5-7.2 PROPERTIES OF ROOTS IN THE s PLANE

The roots of the characteristic equation of a system are either first- or second-order, and each root is associated with a specific transient response mode. The characteristics of the roots, the response modes, and the speci-

fic contours in the s plane are related in a simple way.

5-7.3 First-Order Root: $s = -\frac{1}{T}$

The transient response mode corresponding to this root is $e^{-t/T}$, where T is the time constant of the mode. Lines drawn parallel to the imaginary axis in the s plane are loci of constant T for first-order factors.

5-7.4 Second-Order Root:

$$s = -\zeta \omega_n \pm j\omega_n \sqrt{1 - \zeta^2}$$

The transient response mode corresponding to this root is

$$e^{-\zeta \omega_n t} \cos \omega_d t$$

where

$$(\zeta \omega_n)^{-1} = \text{time constant of envelope of mode}$$

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} = \text{damped frequency of transient oscillation}$$

$$\zeta = \text{damping ratio}$$

$$\omega_n = \text{undamped natural frequency}$$

For the second-order root, s -plane loci can be developed by using the following properties:

$$\text{Re}(s) = -\zeta \omega_n \quad (5-51)$$

$$\text{Im}(s) = \pm \omega_d = \pm \omega_n \sqrt{1 - \zeta^2} \quad (5-52)$$

$$|s| = \omega_n \quad (5-53)$$

$$\angle s = \pm \cos^{-1} \zeta \pm 180^\circ \quad (5-54)$$

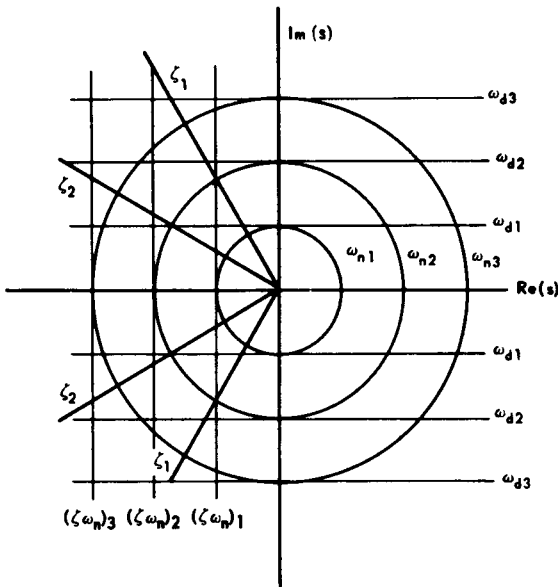


Fig. 5-31 Loci of characteristic parameters of second-order root.

Thus, lines drawn parallel to the imaginary axis are loci of constant-envelope time constant $(\zeta\omega_n)^{-1}$. Lines drawn parallel to the real axis are loci of constant damped frequency of oscillation ω_d . Circles centered at the origin are loci of constant natural frequency ω_n . Radial lines emanating from the origin are loci of constant damping ratio ζ . The various s -plane loci for the second-order factor are shown in Fig. 5-31.

5-7.5 GAIN DETERMINATION IN THE s PLANE

The usual degree-of-stability criterion for determining the gain K from the root locus of a system is:

The dominant roots are adjusted to satisfy a specified damping ratio ζ .

The advantage of the root-locus procedure over the M_p criterion of the frequency-response method becomes evident if a truly dominant pole pair exists. In this case, all other poles are far from the origin, and the transient response of the system is dominated by the transient response mode associated with the dominant pole pair. Thus, the time

behavior of the system becomes immediately evident once the damping-ratio criterion is satisfied. The disadvantage of the root-locus method is that considerable time is consumed in constructing the locus. Adjustment of the gain K for a specified dominant-root damping ratio is best demonstrated by a specific example.

Example. The open-loop function $G(s)$ of a unity-feedback system is

$$G(s) = \frac{K(0.2s + 1)}{s(s + 1)(0.1s + 1)} \quad (5-55)$$

Find the gain K and the closed-loop pole-zero configuration for a dominant-root damping ratio $\zeta = 0.5$.

Solution

(a) The open-loop function is placed in the standard form of the root-locus method as follows:

$$G(s) = 2K \frac{(s + 5)}{s(s + 1)(s + 10)} \quad (5-56)$$

(b) The angle condition is

$$\angle(s + 5) - \angle s - \angle(s + 1) - \angle(s + 10) = -180^\circ \quad (5-57)$$

(c) The magnitude condition is

$$2K \frac{|s + 5|}{|s| |s + 1| |s + 10|} = 1 \quad (5-58)$$

(d) The locus of the roots is constructed from the angle condition by choosing trial points in the s plane and checking back to see whether the angles of the vectors of the open-loop poles and zeros add up according to Eq. (5-57). A curve drawn through the trial points that satisfy this equation is the root locus. To determine the gain K associated with each locus point, Eq. (5-58) is used. The value of s corresponding to a given locus point is substituted into this magnitude equation and the equation is then solved for K . The locus for this problem in the upper half of the s plane is shown in Fig. 5-32.

(e) The line corresponding to $\zeta = 0.5$ ($\angle s = \pm 120^\circ$) is drawn and the intersection with the locus is noted. The intersection occurs at $s = -0.60 \pm j1.04$.

(f) Using the magnitude condition, it is found that $K = 1.38$.

GAIN DETERMINATION

(g) For $K = 1.38$, the location of the real root between $s = -5$ and $s = -10$ is determined by direct application of the magnitude condition. This root lies at $s = -9.85$.

(h) The closed-loop pole-zero configuration has a zero at the open-loop zero $s = -5$.

(i) The factored closed-loop transfer function is

$$W(s) = \frac{2.84 (s + 5)}{(s + 9.85) (s^2 + 1.2s + 1.44)}$$

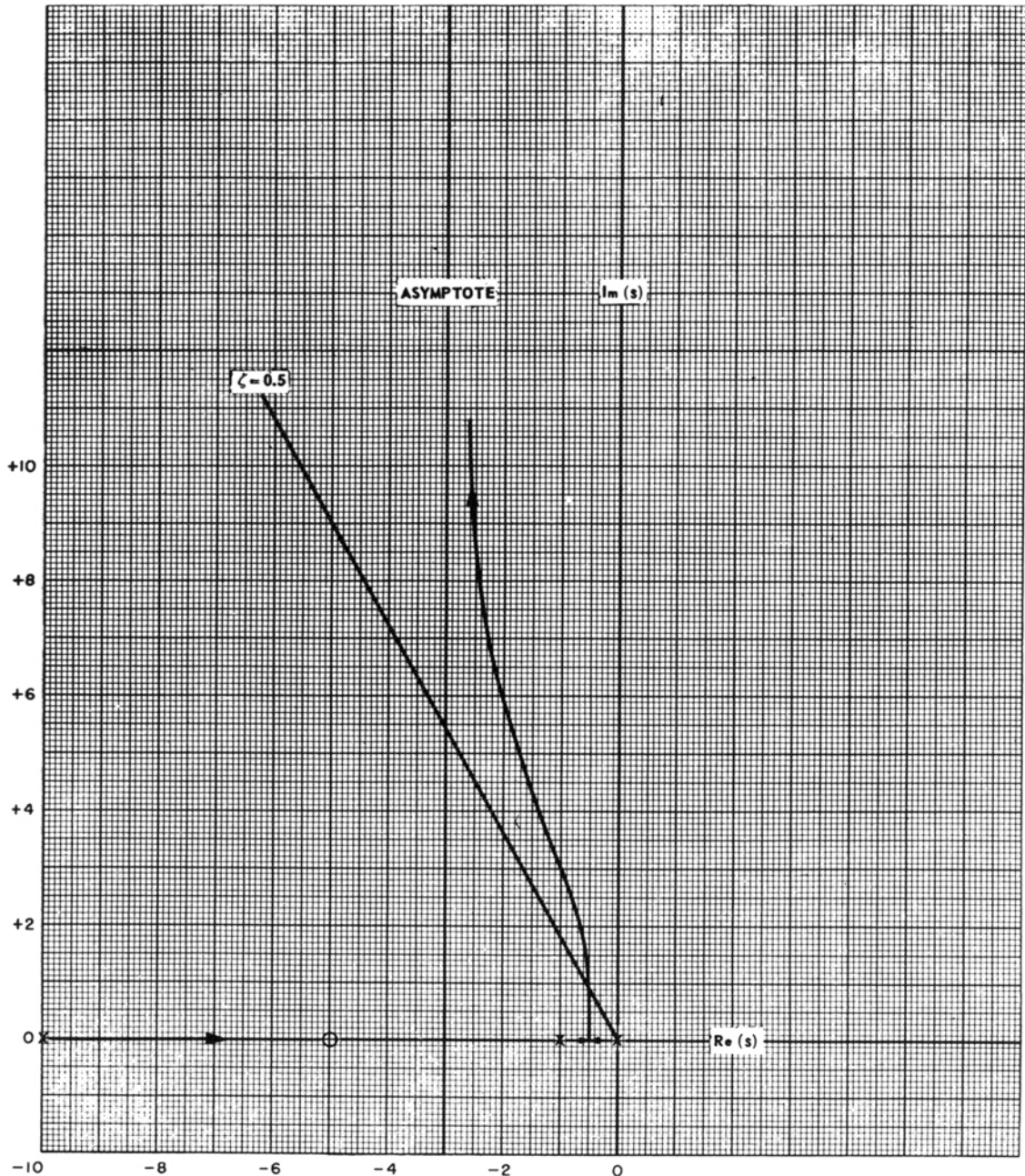


Fig. 5-32 Gain determination from root locus, $G(s) = \frac{K(0.2s + 1)}{s(s + 1)(0.1s + 1)}$.

THEORY

BIBLIOGRAPHY

- 1 G. S. Brown and D. P. Campbell, *Principles of Servomechanisms*, pp. #92-104, John Wiley & Sons, Inc., New York, N. Y., 1948.
- 2 *ibid.*, pp. #146-154.
- 3 *ibid.*, pp. #166-168.
- 4 *ibid.*, pp. #176-189.
- 5 *ibid.*, pp. #236-261.
- 6 J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. #80-81, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.
- 7 *ibid.*, pp. #221-242.
- 8 W. R. Evans, *Control-System Dynamics*, pp. #96-113, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.
- 9 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms*, MIT Radiation Laboratory Series, Vol. 25, pp. #158-186, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.
- 10 W. R. Ahrendt and J. F. Taplin, *Automatic Feedback Control*, McGraw-Hill Book Company, Inc., New York, N. Y., 1951.
- 11 Edited by I. A. Greenwood, Jr., J. V. Holdam, Jr., and D. MacRae, Jr., *Electronic Instruments*, MIT Radiation Laboratory Series, Vol. 21, pp. #215-318, McGraw-Hill Book Company, Inc., New York, N. Y., 1948.
- 12 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. I, pp. #221-244, John Wiley & Sons, Inc., New York, N. Y., 1951.
- 13 *ibid.*, pp. #291-326.
- 14 *ibid.*, pp. #347-350.

CHAPTER 6

COMPENSATION TECHNIQUES*

6-1 INTRODUCTION

Compensation in the general field of servomechanisms refers to the procedures used to modify the dynamic response characteristics of a system by auxiliary means so that it meets performance specifications. Most actual components have a limited dynamic response and so do not respond instantaneously to input variations. Because of the dynamic limitations of physical components, stability problems arise in closed-loop systems, as discussed in Chs. 4 and 5. The requirement of stable operation imposed on all closed-loop systems limits the accuracy that can be obtained with these systems. In order to minimize this inherent limitation, artificial means (compensation techniques) are used to modify the dynamic characteristics. These include the introduction of networks cascaded with the fixed elements in the loop and the addition of auxiliary loops to the system.

The general compensation problem is illustrated by Fig. 6-1. In this figure, $G_f(s)$ represents the response of the fixed elements in the loop which cannot be altered, $H(s)$ represents the response of feedback elements that may be present, and $G_c(s)$ represents the response of compensating elements that are to be adjusted so that the complete system meets the performance specifications. The procedure for designing the system can be outlined as follows [$H(s)$ is assumed to be unity]:

(a) With $G_c(s) = K$, a pure gain (real number), a stability check is made to determine the allowable range of the gain K for stable operation.

(b) Assuming a specified degree of stability, the gain K , is adjusted to meet this requirement.

(c) From the input specifications, the error of the system is found when K is adjusted as in (b), and this error is checked against the error specification.

(d) If the error does not meet specifications, a more complicated form for $G_c(s)$ is introduced. The system is then adjusted to satisfy the specified degree of stability, and the error specification is again checked.

(e) The procedure is continued, trying different or more complicated compensation functions, until the error falls within specifications (that is, if the specifications can be met).

In practice, the forms of compensation normally employed are kept simple. This is due in part to the fact that the fixed elements are usually limited in their range of linear operation, and the introduction of complex compensation functions merely reduces the range over which the linearity assumption applies. In addition, it is found that the theoretical advantages that may accrue with complex compensation are not realized in practice because the theoretical model no longer fits the physical system.

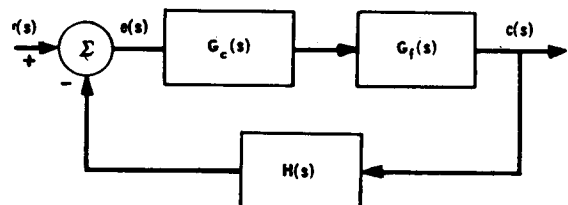


Fig. 6-1 Compensation in a single loop.

*By L. A. Gould

6-2 RESHAPING LOCUS ON GAIN-PHASE PLANE ^(1,2,3,4,8,11)

6-2.1 GENERAL

Because magnitudes and phase angles add when functions are cascaded in the gain-phase plane, this frequency domain is the one most suitable for studying the effects of cascaded compensation functions. The introduction of a compensation function in the loop of a unity-feedback system can be thought of as a method for reshaping the open-loop frequency response $C(j\omega)/E(j\omega)$ to permit a higher gain to be used for the specified degree of stability. If M_p is the degree-of-stability criterion used, the gain can be increased by causing the phase and gain margins of the function $G_c(j\omega)G_f(j\omega)/K$ to increase by a proper choice of $G_c(j\omega)$. To maintain the specified M_p , the M_p contour must be moved *downward* for tangency to occur; this downward motion corresponds to an increase in the open-loop gain of the system (Fig. 6-2). The two most commonly used compensation networks for reshaping the open-loop gain-phase locus are the lag network and the lead network (see Par. 6-6).

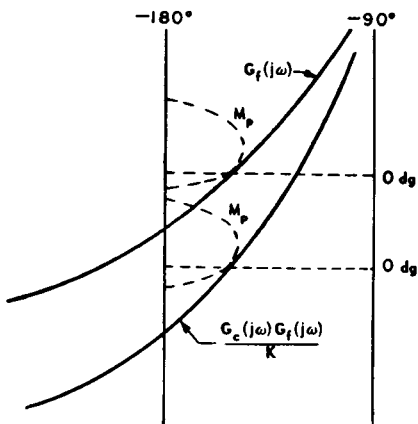


Fig. 6-2 Change in open-loop response produced by compensation illustrating downward motion of M_p contour for gain increase

6-2.2 LAG COMPENSATION

The first-order lag function is

$$G_c(j\omega) = K \frac{T_c j\omega + 1}{\alpha T_c j\omega + 1}; \quad \alpha > 1 \quad (6-1)$$

where K is the real gain factor, T_c is the time constant, and α is an attenuation factor. Lag-function plots for $\alpha = 5, 10, 20$ (with $K = 1$) are shown in Fig. 6-3. These plots are made to a normalized frequency scale for which

$$\beta = T_c \omega \quad (6-2)$$

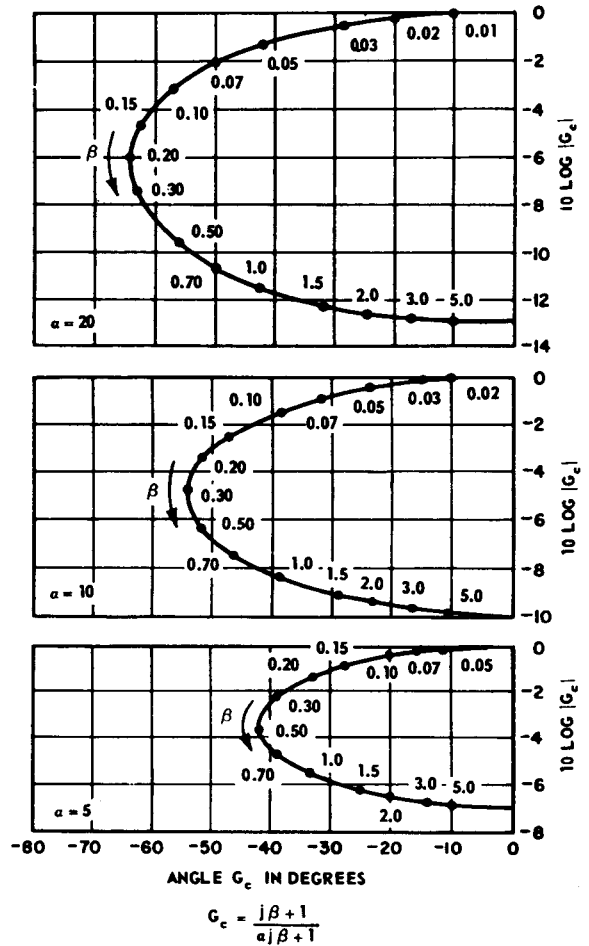


Fig. 6-3 Universal lag functions.

COMPENSATION TECHNIQUES

In using the lag function to reshape an open-loop frequency response, the choice of the attenuation factor is usually governed by the gain increase that is sought. In practice, α 's greater than 20 are rarely used. A good rule of thumb is that the gain increase that can be achieved lies in the range from 0.7α to 0.9α . There are two considerations that limit the choice of the time constant T_c . Since the lag function introduces a negative phase shift, the choice of the time constant should be such that this phase shift does not occur in the region where the open-loop response passes near the $-1+j0$ point. Consequently, the lag function is usually adjusted so that its major phase contribution occurs at low frequencies. This means that T_c cannot be made too small without affecting the stability of the system. On the other hand, if T_c is made too large, the transient response of the system tends to deteriorate as a result of excessive peaking and an abnormally long settling time. A rule of thumb for adjusting the lag function is as follows: Choose T_c so that a phase shift of -5° to -10° is introduced at the uncompensated resonant frequency of the system. The uncompensated resonant frequency is defined as the resonant frequency which is obtained for the specified M_p when $G_c(s) = K$.

The steps involved in adjusting the lag function are the following:

(a) With $G_c(s) = K$, the gain and resonant frequency of the system are found for the specified value of M_p .

(b) Using Fig. 6-3, the value of β is determined for the specified allowable phase shift. Since the region of low phase shift of the lag function occurs for values of $\beta \geq 5$, the phase shift of the lag function in this region is adequately represented by

$$\angle G_c(j\beta) \cong -\left(\frac{\alpha-1}{\alpha}\right)\frac{1}{\beta} \text{ for } \beta \geq 5 \quad (6-3)$$

[†] Symbol \angle denotes "angle"

Thus, the phase angle varies inversely as β .

(c) If ω_{R1} is the uncompensated resonant frequency and β_ϕ is the value of β corresponding to the specified allowable phase shift of the lag function, then, from Eq. (6-2),

$$T_c = \frac{\beta_\phi}{\omega_{R1}} \quad (6-4)$$

(d) Since the scale ratio between β and ω is fixed by Eq. (6-4), the magnitude and phase-angle contribution of the lag function to the $G_r(j\omega)$ function at each frequency can be determined from the universal curves of Fig. 6-3.

(e) After the lag function has been added to the fixed-element response $G_r(j\omega)$, the gain K is determined from the specified M_p criterion.

Example. The transfer function of the fixed elements of a unity-feedback system is given by

$$G_r(s) = \frac{1}{s(0.3s+1)(0.1s+1)} \quad (6-5)$$

A lag function with $\alpha = 10$ is used to compensate the system. The lag function is to contribute -5° of phase shift at the uncompensated resonant frequency. Design the compensation when $10 \log_{10} M_p = 1.5$ dg.

Solution. The frequency response $G_r(j\omega)$ is plotted in Fig. 6-4 as Curve A. For the specified M_p (see construction), the point of tangency of the M_p contour with Curve A occurs at the point where $\omega_{R1} = 2.4$ rad/sec; from the displacement downward of the M_p contour by 4.1 dg (i.e., $10 \log K = 4.1$) one gets $K = 2.57$ for the uncompensated system when $G_c(j\omega) = K$. From the $\alpha = 10$ -plot of Fig. 6-3, -5° of phase shift occurs at $\beta_\phi = 10$. Therefore,

$$T_c = \frac{\beta_\phi}{\omega_{R1}} = \frac{10}{2.4} = 4.17 \text{ seconds} \quad (6-6)$$

The scale change from ω to β is, therefore,

$$\beta = 4.17\omega \quad (6-7)$$

THEORY

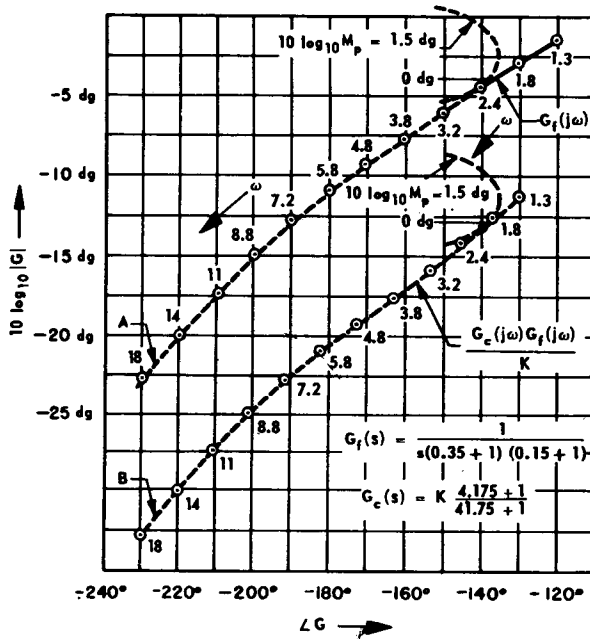


Fig. 6-4 Lag-compensation procedure.

With this scale change, the universal lag function for $\alpha = 10$ is used to add the magnitude and phase angle of $G_c(j\omega)/K$ to the fixed-element response $G_f(j\omega)$. The composite $G_c(j\omega)G_f(j\omega)/K$ function appears as Curve B in Fig. 6-4. Using the specified M_p criterion, the resonant frequency of the compensated system is 2.0 rad/sec and K has been increased to a value $K = 20.4$. Thus, the use of the lag function with $\alpha = 10$ has allowed an increase in gain by a factor of 7.6. The resonant frequency has been decreased by 17 percent.

6-2.3 LEAD COMPENSATION

The first-order lead function is

$$G_c(j\omega) = K \frac{\alpha T_c j\omega + 1}{T_c j\omega + 1}; \quad \alpha > 1 \quad (6-8)$$

where K is the gain, T_c is the time constant, and α is an attenuation factor. Lead-function plots for $\alpha = 5, 10$, and 20 are shown in Fig.

6-5. These plots utilize a normalized frequency scale for which

$$\beta = T_c \omega \quad (6-9)$$

In using the lead function to reshape an open-loop frequency response, advantage is taken of the fact that the lead function exhibits positive phase shift. Thus, by adjusting the time constant T_c , it is possible to add positive phase angles to the fixed-element response $G_f(j\omega)$ in a region where the negative phase shift of the fixed elements is too great to secure an M_p -contour tangency. Hence, the lead func-

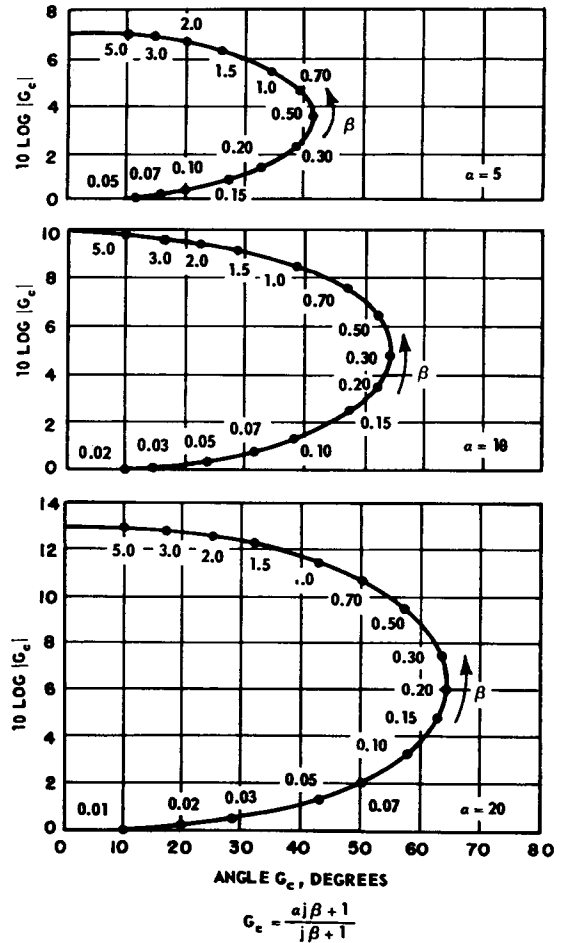


Fig. 6-5 Universal lead functions.

COMPENSATION TECHNIQUES

tion can decrease the effective negative phase shift of the composite $G_c(j\omega)G_f(j\omega)/K$ function, allowing an M_p tangency to occur for a larger value of gain K . In many cases, an increase in the resonant frequency of the system is obtained when lead compensation is used.

Due to the tendency of lead compensation to increase the bandwidth of a system, it is found that the system is more sensitive to noise, and its linear range of operation is restricted. Thus, in practical situations, the attenuation factor α used usually does not exceed 20. The adjustment of the time constant T_c is a matter of trial and error. An outline of the trial-and-error procedure is given below:

(a) The phase-angle difference ψ between the "nose" of the M_p contour and the -180° line of the gain-phase plane is given by

$$\psi = \sin^{-1} \left(\frac{1}{M_p} \right) \quad (6-10)$$

The maximum phase shift ϕ_m introduced by the lead function is

$$\phi_m = \sin^{-1} \left(\frac{\alpha - 1}{\alpha + 1} \right) \quad (6-11)$$

(b) The first-trial choice of the lead-function time constant T_c is found by determining the frequency at which the following angle relation holds:

$$\angle G_f(j\omega_1) = -180^\circ + \psi - \phi_m \quad (6-12)$$

The frequency ω_1 which satisfies Eq. (6-12) be found directly from the gain-phase plot of $G_f(j\omega)$. Then, the first choice of T_c is given by

$$T_{c1} = \frac{\beta_m}{\omega_1} \quad (6-13)$$

where β_m is the frequency at which the maximum phase shift (ϕ_m) occurs on the normalized lead-function curve of Fig. 6-5. The frequency β_m can be found from the relation

$$\beta_m = \frac{1}{\sqrt{\alpha}} \quad (6-14)$$

(c) Since the scale ratio between ω and the normalized frequency β is fixed by Eq. (6-14), the magnitude and phase-angle contribution of the lead function to the $G_f(j\omega)$ function at

each frequency can be determined from the universal curves of Fig. 6-5.

(d) After the lead function has been added to the fixed-element response $G_f(j\omega)$, the gain K is determined from the specified M_p criterion.

(e) The first-trial choice of the lead-function time constant T_c usually determines a closed-loop resonant frequency that is close to the maximum obtainable frequency with the given attenuation factor α .

However, the open-loop gain K is not necessarily a maximum. Therefore, if gain increase is the objective, additional trials must be made. The additional trials usually involve the choice of trial values of T_c that are smaller than the initial choice. A rule of thumb is that the time constant T_c which maximizes K is approximately one-half to one-third the initial-trial value.

Example. The transfer function of the fixed elements of a unity-feedback system is given by Eq. (6-5). A lead function with $\alpha = 10$ is used to compensate the system. Design the compensation when $X \log_{10} M_p = 1.5$ dg.

Solution. The frequency response $G_f(j\omega)$ is plotted in Fig. 6-6 (Curve A). From the problem specifications and Eqs. (6-10), (6-11),

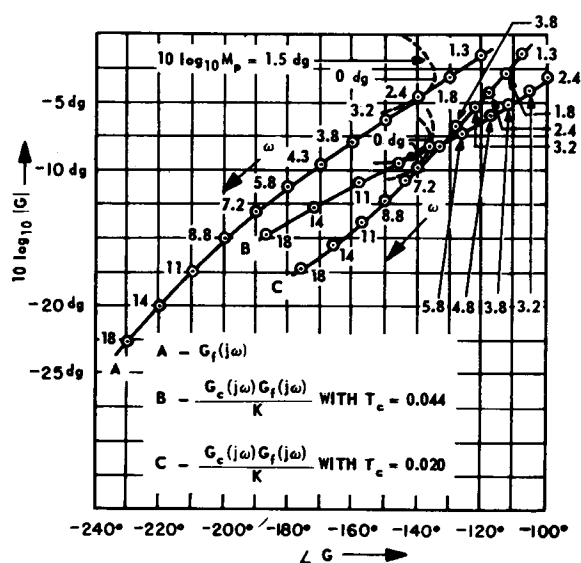


Fig. 6-6 Lead-compensation procedure.

THEORY

and (6-14), we find that $\psi = 45^\circ$, $\phi_m = 54.9^\circ$, and $\beta_m = 0.316$. Eq. (6-12) yields

$$\angle G_f(j\omega_1) = -189.9^\circ \quad (6-15)$$

Using this result and Curve A of Fig. 6-6, we find that $\omega_1 = 7.2$ rad/sec. Then, from Eq. (6-13), the initial choice for the lead-function time constant is $T_c = 0.044$ sec. The scale change from ω to β is therefore given by

$$\beta = 0.044\omega \quad (6-16)$$

With this scale change, the universal lead function for $\alpha = 10$ is used to add the magnitude and phase angle of $G_c(j\omega)/K$ to the fixed-element response $G_f(j\omega)$. The composite $G_c(j\omega)G_f(j\omega)/K$ function appears as Curve B in Fig. 6-6. Using the specified M_p criterion, the resonant frequency of the uncompensated system (Curve A) is $\omega_{R_1} = 2.4$ rad/sec and

the gain of the uncompensated system with $G_c(j\omega) = K$ is $K = 2.57$. The resonant frequency of the compensated system (Curve B) is $\omega_{R_2} = 8.2$ rad/sec and the gain of the compensated system is $K = 6.68$. Thus, the resonant frequency of the system has been increased by a factor of 3.4 and the gain by a factor of 2.6 through lead compensation. If maximum gain is sought, the lead time constant must be reduced. By trial and error, we find that with $T_c = 0.020$ sec, $\omega_R = 5.6$ rad/sec and $K = 8.65$. The construction for maximum gain is shown as Curve C in Fig. 6-6. Thus, for maximum gain adjustment, lead compensation increases the resonant frequency by a factor of 2.3 and the gain by a factor of 3.4. Note that the time constant for maximum gain is 0.45 times the initial-trial choice. The results of this example are listed in Table 6-1.

TABLE 6-1 RESULTS OF LEAD COMPENSATION EXAMPLE

| Compensation Adjustment | T_c (sec) | ω_R (rad/sec) | K |
|---|-------------|----------------------|------|
| Initial trial; maximum resonant frequency | 0.044 | 8.2 | 6.68 |
| Final trial; maximum gain | 0.020 | 5.6 | 8.65 |
| No compensation | | 2.4 | 2.57 |

6-3 PHASE-MARGIN AND ASYMPTOTIC METHODS ^(3,4,8)

6-3.1 GENERAL

A rough picture of the effect of compensation can be obtained if the magnitude asymptotes are used in conjunction with the phase-margin criterion for the degree of stability (see Par. 5-6). Using a 45° phase-margin criterion, the asymptotic method gives good results provided there are no low-damping-ratio quadratic factors in the open-loop transfer function $C(j\omega)/E(j\omega)$. If the 45° phase-margin criterion is assumed, then the asymptote crossover frequency ω_c (defined in Par.

5-1) usually occurs in a region where the slope of the asymptote is -10 dg/decade.

The phase-margin criterion can be used as an approximation to the M_p criterion, or it can be used as an independent degree-of-stability criterion. If it is used independently, the separate magnitude vs frequency and phase-angle vs frequency plots are employed. If the phase-margin criterion is used to approximate the M_p criterion, one might just as well use the asymptotic curve as an approximation to the true magnitude curve.

The rules of thumb established in Par. 6-2 for adjusting the compensation functions can be applied in an identical manner when working with the separate magnitude and phase angle plots except that the asymptotic-cross-over frequency ω_c or the magnitude-crossover frequency ω_{cm} will replace the resonant frequency ω_R where necessary. The universal compensation-function curves of Figs. 6-3 and 6-5 can be used when working with the separate magnitude and phase-angle curves.

The steps involved in adjusting the compensation functions when using the phase-margin criterion and the separate response curves are outlined below.

6-3.2 LAG COMPENSATION

(a) With $G_c(s) = K$, the gain and magnitude-crossover frequency (or asymptote-crossover frequency) are found for the specified phase margin.

(b) Using the universal lag-function curve (Fig. 6-3), the normalized frequency β_ϕ (at which the allowable negative phase shift of the lag function occurs) is determined.

(c) The lag-function time constant is found from the relation

$$T_r = \frac{\beta_\phi}{\omega_{cm_1}} \quad (6-17)$$

where ω_{cm_1} is the magnitude-crossover frequency of the uncompensated system, or, alternatively, one can use the relation

$$T_r = \frac{\beta_\phi}{\omega_{c_1}} \quad (6-18)$$

where ω_{c_1} is the asymptote-crossover frequency of the uncompensated system.

(c) Using the scale ratio between β and ω determined by Eq. (6-17) or (6-18), the magnitude (or asymptote) and phase-angle contribution of the lag function to the $G_r(j\omega)$ function at each frequency can be determined.

(d) After the lag function has been added to the fixed-element response $G_r(j\omega)$, the gain K is determined from the specified phase margin.

6-3.3 LEAD COMPENSATION

(a) The frequency ω_1 which satisfies Eq. (6-13) is found directly from the $\angle G_r(j\omega)$ curve.

(b) If β_m is the normalized frequency at which the maximum phase shift ϕ_m for the lead function occurs (see Fig. 6-5), then the first trial choice of the lead-compensation time constant is

$$T_{c_1} = \frac{\beta_m}{\omega_1} \quad (6-19)$$

(c) Using the scale ratio between β_m and ω_1 determined by Eq. (6-19), the magnitude (or asymptote) and phase-angle contribution of the lead function to the $G_r(j\omega)$ function at each frequency can be determined.

(d) After the lead function has been added to the fixed-element response $G_r(j\omega)$, the gain K is determined from the specified phase margin.

(e) Further trial values of the lead-function time constant T_c are tried until the gain K or the magnitude- (or asymptote-) crossover frequency has been maximized.

Example. The transfer function of the fixed elements of a unity-feedback system is given by Eq. (6-5). The 45° phase-margin criterion is to be used to adjust the degree of stability of the system.

(a) Lag compensation with $\alpha = 10$ is to be used to improve performance. The allowable negative phase shift that the lag function contributes at the magnitude- (or asymptote-) crossover frequency of the uncompensated system shall be 5° . Design the compensation.

(b) Lead compensation with $\alpha = 10$ is to be used to improve performance. Design the compensation.

Solution.

(a) The magnitude and asymptote curves of $G_r(j\omega)$ are drawn in Fig. 6-7 and the phase angle curve is drawn in Fig. 6-8. For a 45° phase margin, the magnitude-crossover frequency $\omega_{cm} = 2.08$ rad/sec and the corresponding gain is 2.34. The asymptote-crossover frequency $\omega_c = 2.08$ rad/sec and the corresponding gain is 2.04. (Compare with $\omega_R = 2.4$ rad/sec and $K = 2.57$ for $10 \log_{10} M_p = 1.5$ dg.) For 5° of allowable negative

THEORY

phase shift, and $\alpha = 10$, $\beta_\phi = 10$ from Fig. 6-3. Therefore, using Eqs. (6-17) or (6-18), $T_c = 4.8$ sec. The scale change from β to ω is given by

$$\beta = 4.8\omega \quad (6-20)$$

The composite $G_c(j\omega)G_f(j\omega)/K$ magnitude

and asymptote curves for the lag-compensated system are drawn in Fig. 6-7, and the composite phase-angle curve is drawn in Fig. 6-8. Using the 45° phase margin, the magnitude-crossover frequency $\omega_{cm} = 1.74$ rad/sec, and the corresponding gain is 20. The asymptote-crossover frequency $\omega_c = 1.74$ rad/sec and the corresponding gain is 17.8.

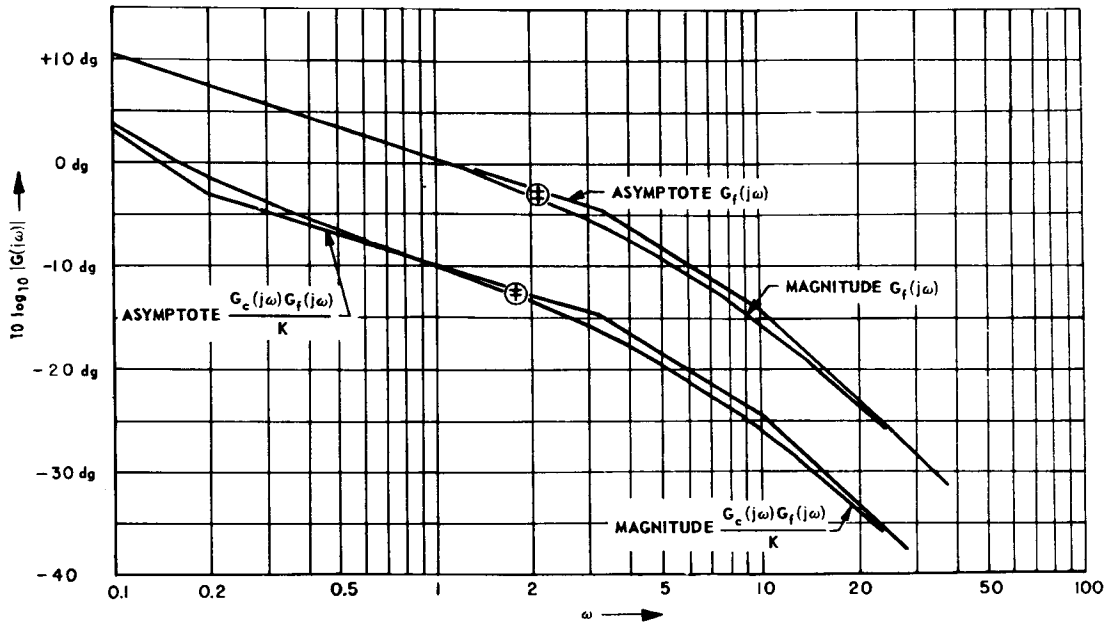


Fig. 6-7 Magnitude curves for lag-compensation procedure employing phase margin.

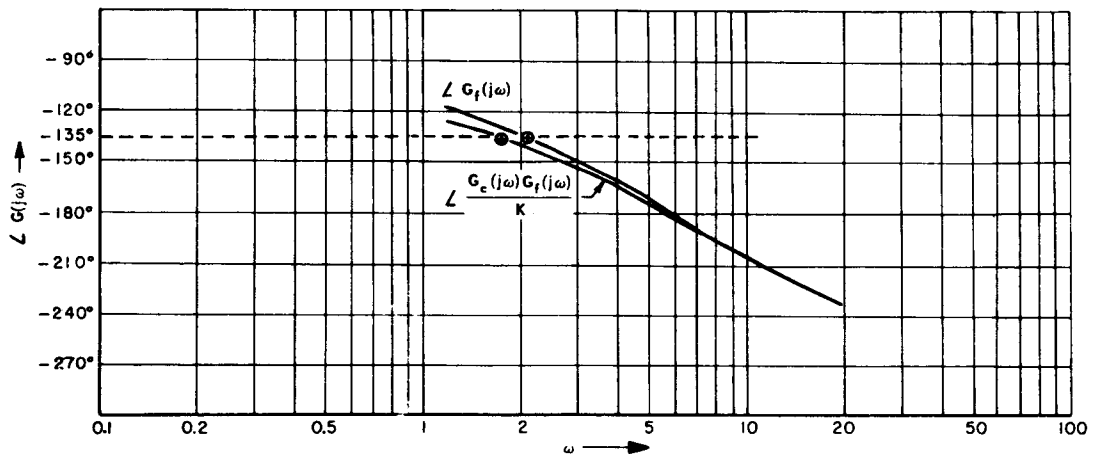


Fig. 6-8 Phase curves for lag-compensation procedure employing phase margin.

COMPENSATION TECHNIQUES

(b) The magnitude and asymptote curves of $G_f(j\omega)$ are drawn in Fig. 6-9 and the phase angle curve is drawn in Fig. 6-10. Using the problem specifications and Eqs. (6-11) and (6-14), $\psi = 45^\circ$, $\phi_m = 54.9^\circ$ and $\beta_m = 0.316$. Eq. (6-12) yields

$$\angle G_f(j\omega_1) = -189.9^\circ \quad (6-21)$$

Using this result and the $G_f(j\omega)$ phase-angle

curve (Fig. 6-10), $\omega_1 = 7.2$ rad/sec. From Eq. (6-19), the initial-trial time constant $T_{c1} = 0.044$ sec and the scale change from β to ω is given by

$$\beta = 0.044\omega \quad (6-22)$$

The composite $G_c(j\omega)G_f(j\omega)/K$ magnitude and asymptote curves for the first-trial lead-compensated system are drawn in Fig. 6-9

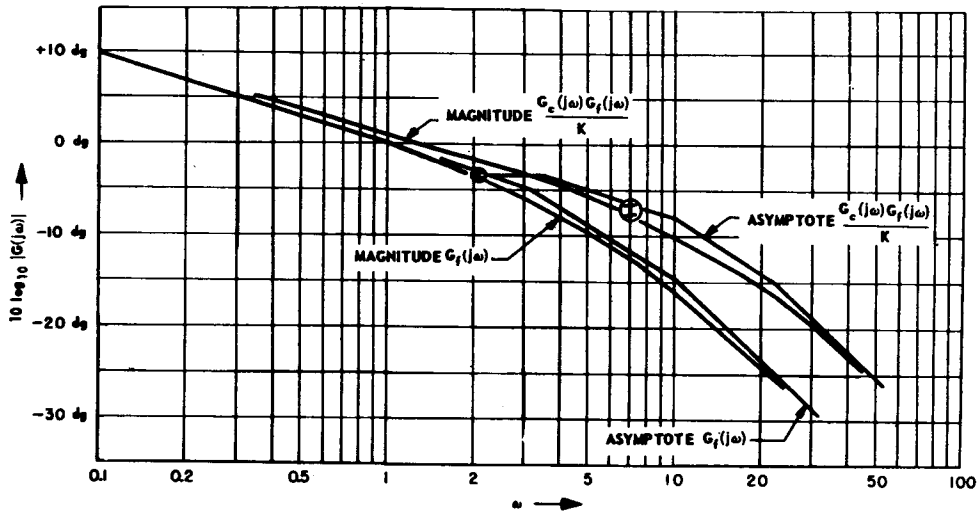


Fig. 6-9 Magnitude curves for lead-compensation procedure employing phase margin.

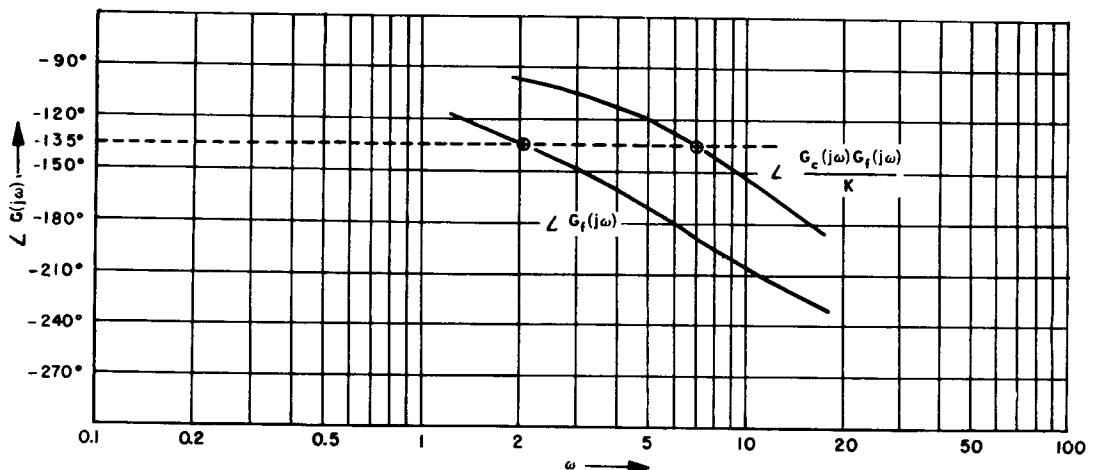


Fig. 6-10 Phase curves for lead-compensation procedure employing phase margin.

THEORY

and the corresponding phase-angle curve is drawn in Fig. 6-10. Using the 45° phase-margin criterion for the lead compensated system, the magnitude-crossover frequency $\omega_{cm} = 7.0$

rad/sec, and the corresponding gain is 6.02. The asymptote-crossover frequency $\omega_c = 7.0$ rad/sec and the corresponding gain is 4.46. The results of this example are summarized in Table 6-2.

TABLE 6-2 RESULTS OF COMPENSATION USING 45° PHASE MARGIN

| Compensation Adjustment | ω_{cm} | K for ω_{cm} | ω_c | K for ω_c | Factor of Gain Increase | Factor of ω_c or ω_{cm} Change |
|--|---------------|-----------------------|------------|--------------------|-------------------------|--|
| No compensation; true magnitude used | 2.08 | 2.34 | | | | |
| No compensation; asymptote used | | | 2.08 | 2.04 | | |
| Lag compensation; true magnitude used | 1.74 | 20 | | | 8.6 | 0.84 |
| Lag compensation; asymptote used | | | 1.74 | 17.8 | 8.7 | 0.84 |
| Lead compensation; true magnitude used | 7.0 | 6.02 | | | 2.6 | 3.4 |
| Lead compensation; asymptote used | | | 7.0 | 4.46 | 2.2 | 3.4 |

6-4 FEEDBACK OR PARALLEL COMPENSATION ^(2,3,4)

The cascade type of compensation discussed in Pars. 6-2 and 6-3 has the disadvantage that the compensation adjustment is sensitive to changes in the parameters of the fixed elements due to non-linear behavior of the system. When feedback compensation is used, on the other hand, the compensation adjustment is much less sensitive to fixed-element parameter variation provided the loop gain is high. In addition, the networks used in feedback compensation are usually simpler in form than the corresponding cascade networks. However, the necessity for high loop gain (at

least a gain of 10 at the break frequencies of the feedback networks) generally requires a more complicated and expensive system.

The procedure used in designing feedback compensation networks employs a combination of the gain-phase plane and the asymptotic-magnitude presentations. The basic principles involved in the design of feedback compensation networks can be clarified by a study of Fig. 6-11. Here a feedback function $H_c(s)$ is used to modify the characteristics of the fixed elements $G_f(s)$ and a cascade function $G_c(s)$ is provided to aid in adjusting the

COMPENSATION TECHNIQUES

performance of the major loop. In most cases, the cascade function $G_c(s)$ is a pure gain K which serves to adjust the degree of stability of the system. The burden of reshaping the $G_c(s)$ function is placed on the feedback compensation function $H_c(s)$. In addition, the feedback function is usually provided with a gain adjustment to permit the setting of the degree of stability of the minor loop.

The general configuration of Fig. 6-11 can be redrawn and placed in the cascade form of Fig. 6-12. Here,

$$G'_r(s) = \frac{G_r(s)}{1 + G_r(s)H_r(s)} \quad (6-23)$$

Thus, theoretically, cascade compensation and feedback compensation are equivalent. In practice, feedback compensation is more flexible, and the resulting system is less sensitive to component parameter variations.

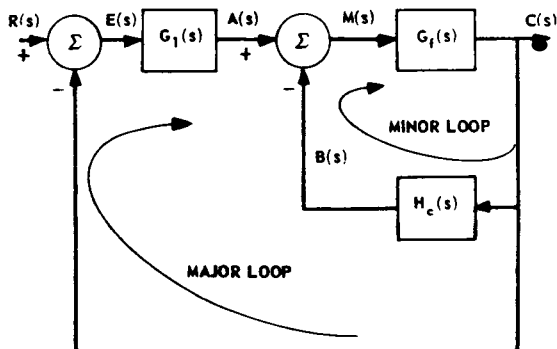


Fig. 6-11 General feedback-compensation configuration.

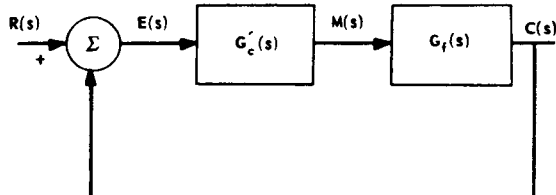


Fig. 6-12 Cascade equivalent of feedback-compensation configuration.

The procedure for adjusting the feedback compensation is best understood by examining the asymptotic behavior of the minor loop. If

$$G_f(j\omega)H_f(j\omega) \gg 1,$$

then

$$\frac{C(j\omega)}{A(j\omega)} = \frac{G_f(j\omega)}{1 + G_f(j\omega)H_r(j\omega)} \cong \frac{1}{H_r(j\omega)} \quad (6-24)$$

If

$$G_f(j\omega)H_c(j\omega) \ll 1,$$

then

$$\frac{C(j\omega)}{A(j\omega)} \cong G_f(j\omega) \quad (6-25)$$

Thus, in the frequency ranges where the open-minor-loop frequency-response magnitude $G_f(j\omega)H_c(j\omega)$ is very large, the closed-minor-loop response $C(j\omega)/A(j\omega)$ behaves like the reciprocal of the feedback function $H_c(j\omega)$. When the open-minor-loop-response magnitude is very small, the closed-minor-loop response behaves like the fixed elements response function $G_f(s)$. Thus, the frequency scale can be divided into several regions based on the magnitude of the open-minor-loop response. Whenever the magnitude of this response $G_f(j\omega)H_c(j\omega)$ or $B(j\omega)/M(j\omega)$ is greater than unity, the asymptotes of the closed-minor-loop response coincide with the asymptotes of the reciprocal of the feedback function $H_c(j\omega)$. Whenever the magnitude of the $B(j\omega)/M(j\omega)$ function is less than unity, the closed-minor-loop response asymptotes coincide with the asymptotes of the fixed-element response $G_f(j\omega)$.

Usually, the feedback function $H_c(j\omega)$ is so chosen that the frequency scale is divided into three ranges. These ranges are:

$$0 < \omega < \omega_l \quad \text{when} \quad |G_f(j\omega)H_c(j\omega)| < 1 \quad (6-26)$$

$$\omega_l < \omega < \omega_u \quad \text{when} \quad G_f(j\omega)H_r(j\omega) > 1 \quad (6-27)$$

$$\omega_n < \omega < \infty \quad \text{when} \quad G_f(j\omega)H_r(j\omega) < 1 \quad (6-28)$$

THEORY

and ω_l is the *lower*-frequency boundary and ω_u is the *upper*-frequency boundary. Corresponding to the three frequency ranges fixed by the magnitude of the open-minor-loop response, the closed-minor-loop response asymptotes are defined by

$$\left| \frac{C(j\omega)}{A(j\omega)} \right| \cong \left| G_r(j\omega) \right| \quad \text{for } 0 < \omega < \omega_l \quad (6-29)$$

$$\left| \frac{C(j\omega)}{A(j\omega)} \right| \cong \left| \frac{1}{H_c(j\omega)} \right| \quad \text{for } \omega_l < \omega < \omega_u \quad (6-30)$$

$$\left| \frac{C(j\omega)}{A(j\omega)} \right| \cong \left| G_r(j\omega) \right| \quad \text{for } \omega_u < \omega < \infty \quad (6-31)$$

The procedure for adjusting the feedback function $H_c(s)$ and the cascade gain $G_c(s) = K$ can thus be roughed out using asymptotic pictures of the various responses and then carried out in detail using the gain-phase plane. The aim of the asymptotic sketches is to examine the form of the closed-minor-loop response as the feedback compensation is adjusted. Examination of the asymptotes of typical cascade compensation arrangements can also serve as a guide to the shaping of the closed-minor-loop response. The desirable properties of the closed-minor-loop response can be expressed in terms of the properties that are desirable for any open-major-loop function; namely, high gain at low frequencies ($\omega < \omega_l$), a stable shape relative to the $-1 + j0$ point, and low gain at high frequencies ($\omega > \omega_u$).

Since there are usually several parameters to adjust in the feedback compensation procedure, the process of adjustment is one of trial and error guided by the asymptotic sketches. The details of the procedure are best demonstrated by an example.

Example. The transfer function of the fixed elements of a system is given by

$$G_r(s) = \frac{1}{s(0.3s + 1)(0.1s + 1)} \quad (6-32)$$

Feedback compensation is to be used to improve the performance of the system in conjunction with a pure gain cascaded with the

minor loop. The transfer function of the feedback elements is given by

$$H_c(s) = \frac{K_c s^2}{T_c s + 1} \quad (6-33)$$

Note that this transfer function can be realized in a position-control system by a tachometer cascaded with a single-stage high-pass RC filter. The cascade compensation being a pure gain, its transfer function is given by

$$C_c(s) = K \quad (6-34)$$

Design the compensation for a 45° major-loop phase margin.

Solution. The open-minor-loop transfer function is

$$\frac{B(s)}{M(s)} = \frac{G_r(s)H_c(s)}{K_c s} = \frac{K_c s}{(T_c s + 1)(0.3s + 1)(0.1s + 1)} \quad (6-35)$$

A plot of the asymptotes of $G_c(j\omega)H_c(j\omega)/K_c$, using the techniques for plotting asymptotic magnitude curves of Par. 5-3.3, is shown in Fig. 6-13 for $T_c = 2$ sec as a trial guess. The adjustment of K_c controls the degree of stability of the closed-minor-loop response. If K_c is made too large (e.g., greater than 50), the closed-minor-loop response will have a quadratic factor with a very low damping ratio, making it difficult to obtain a high-gain open-major-loop response. Anticipating this behavior, a value of $K_c = 10$ is not unreasonable. With $K_c = 10$, the inequality of Eq. (6-27) is satisfied for the portions of the open-minor-loop asymptotes above the -10 dg line. When K_c is set equal to 10, the 0-dg line for the open-minor-loop response $G_r(j\omega)H_c(j\omega)$ is that shown dashed in Fig. 6-13. This line defines the frequency boundaries $\omega_l = 0.1$ rad/sec and $\omega_u = 13$ rad/sec. The closed-minor-loop response asymptotes can then be drawn with the aid of the approximations to $|C(j\omega)/A(j\omega)|$ given in Eqs. (6-29) to (6-31). In the frequency range from ω_l to $\omega = 1/T_c$ there is a -20 dg/dec contribution to $|1/H_c(j\omega)|$ from

COMPENSATION TECHNIQUES

the factor s^2 whereas the $(T_c s + 1)$ factor makes no contribution (i.e., it contributes 0 dg/dec). In the range from $\omega = 1/T_c$ to ω_u , both factors contribute, the s^2 factor contributing -20 dg/dec and the $(T_c s + 1)$ factor $+10$ dg/dec. The closed-minor-loop asymptotes are shown in Fig. 6-14. From an examination of the resultant asymptotic curve, several points may be noted. The combination of the breaks at ω_l and $1/T_c$ appears as a cascade lag-compensation effect which is a desirable open-major-loop-response property. In further trials, attempts may be made to broaden the -20 dg/dec slope region bounded by these breaks and to move the region to a higher frequency range. The break at ω_u is from a slope of -10 dg/dec to a slope of -30 dg/dec which is characteristic of a quadratic factor in the open-major-loop response. If this factor has a low damping ratio, high-gain stabilization of the major loop may be difficult. Thus, the first trial choices of T_c and K_c produce a set of open-major-loop asymptotes which appear reasonable; however the adequacy of the choices must be verified.

At this point, the Nichols chart (Fig. 5-20) is used with a gain-phase plot of $G_f(j\omega)H_c(j\omega)$ to determine the closed-minor-loop response $C(j\omega)/A(j\omega)$ (see Par. 5-4 for the use of the Nichols chart with non-unity-feedback loops).

As a result of the application of the gain-phase plane construction, the true magnitude curve of $C(j\omega)/A(j\omega)$ is shown in Fig. 6-14 can be obtained. The corresponding phase angle curve appears in Fig. 6-15. The shape of the true magnitude curve shows no severe resonance effects so that acceptable closed-loop performance may be expected. Using the 45° phase-margin criterion to adjust the cascade compensation $G_c(s) = K$, the magnitude-crossover frequency $\omega_{cm} = 8.6$ rad/sec and the corresponding gain $K = 40$. Thus, the performance is quite good. (Compare the results of cascade compensation for this same system in Pars. 6-2 and 6-3.) The only drawback to the design is that the equivalent cascade lag effect is at a fairly low frequency. This would produce somewhat excessive peaking in the transient response of the system which would be followed by a long transient tail. Improvement in performance could be achieved by further trial, e.g., by decreasing the feedback compensation time constant T_c , and attempting to increase the minor-loop gain K_c . The resultant system would then have a more acceptable transient behavior, but the magnitude-crossover frequency ω_{cm} and the major-loop gain K might be reduced. However, only further trial-and-error analysis would show what actually occurs.

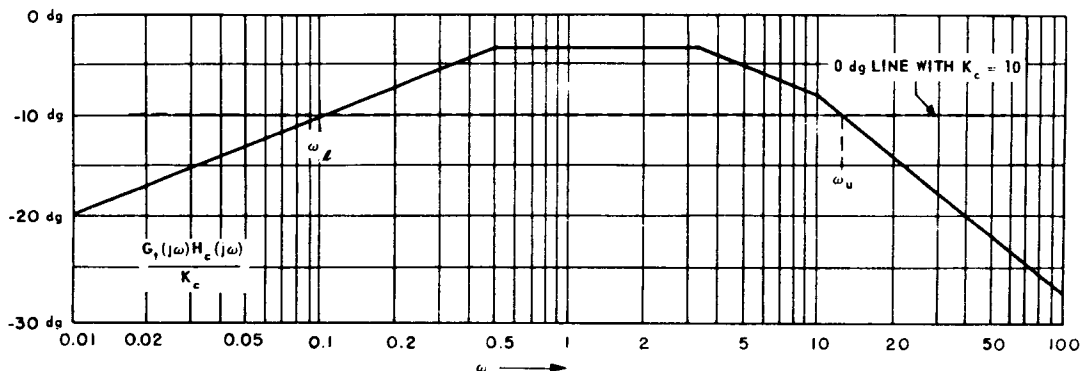


Fig. 6-13 Open-minor-loop asymptote for feedback compensation procedure.

THEORY

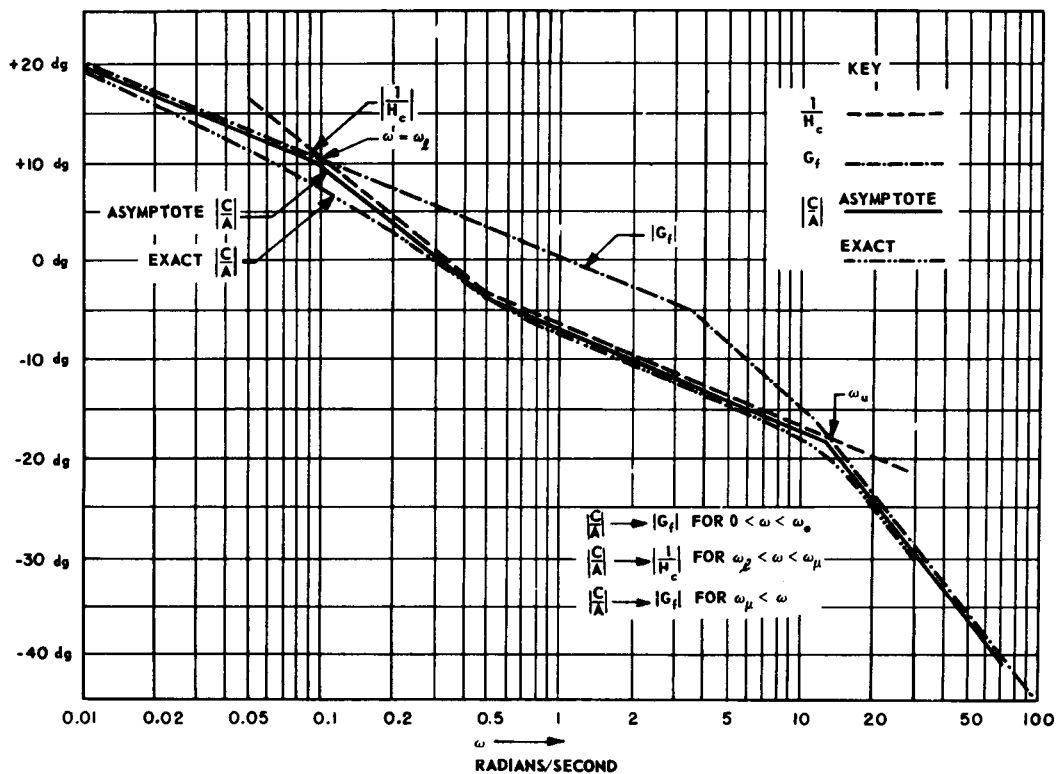


Fig. 6-14 Closed-minor-loop magnitude for feedback compensation procedure.

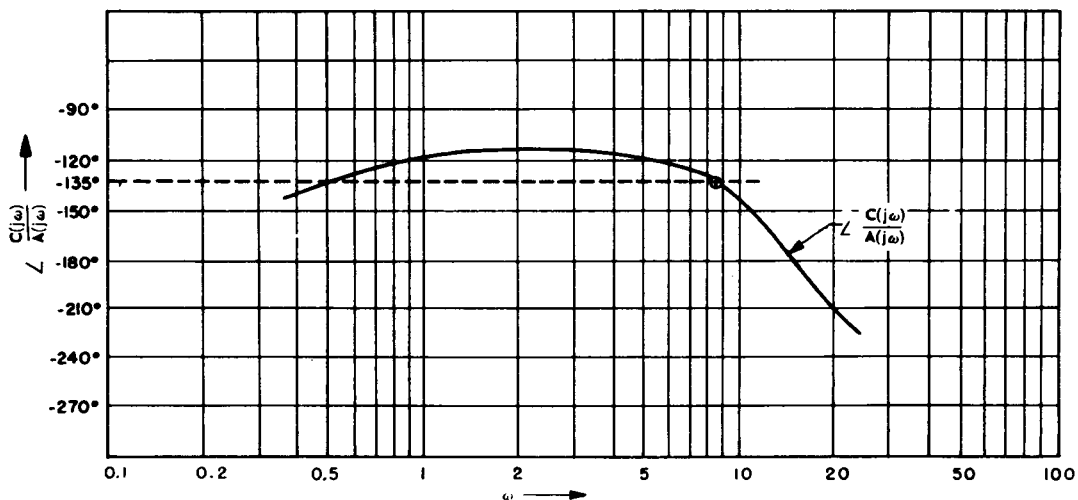


Fig. 6-15 Closed-minor-loop phase angle for feedback compensation procedure.

6-5 ALTERNATIVE DESIGN METHODS^(6,7,9,10,12,13)

The primary advantage of design in the frequency domain is the rapidity with which the procedure can be carried out. The disadvantage is the difficulty involved in visualizing the time-domain behavior corresponding to a given frequency-domain design. In practice, the relationship between frequency response and transient response is considered by many workers to be rather nebulous. Theoretically, however, frequency-domain and time-domain descriptions are entirely equivalent, although the actual process of translating from one description to the other may be quite laborious in spite of the fact that many approximations have been established to guide the designer in relating frequency response with transient response (see Par. 7-1). One very important stumbling block arises because most direct specifications of system performance are given in terms of the transient response of the system to a step or ramp input. This type of specification is just as artificial as that given in terms of the response of a system to a sinusoidal input since the true inputs of most systems are neither steps, nor ramps, nor sinusoids. Nevertheless, it is the transient response of a system that is most frequently specified because this type of response is the easiest to visualize and the quickest to verify experimentally.

In order to circumvent some of the conceptual difficulties involved in frequency-domain design, methods of time-domain design have been advanced. Most of these methods utilize the open-loop and closed-loop pole-zero configurations of the system and thus involve features of both the time and frequency domains. The facility with which these methods can be used depends almost exclusively on the feature of having an analytical description of the open-loop pole-zero configuration as a starting point. Thus, the methods require that any experimental test data be approximated by analytical functions. This requirement does not apply to the frequency-domain methods that receive major emphasis in Ch. 5 and Pars. 6-2 through 6-4. In addition, the

graphical procedures discussed in Pars. 3-6 and 7-1 enable the designer to work entirely with experimental data, going back and forth between time and frequency domains without ever having to deal with analytical descriptions. Since the time-domain synthesis methods usually end up with a closed-loop pole-zero configuration, additional labor is necessary to extract the actual plots of transient response and frequency response in order to verify whether performance specifications have been met. On the other hand, in a frequency-domain design, the only additional labor involved is that of determining the transient response (usually by the methods of Par. 3-6), the frequency response being directly available at the end of the design procedure. Thus far, the time-domain procedures that have been developed are most successful for analysis but are quite time-consuming and laborious for synthesis. Actually, most of the current time-domain "synthesis" procedures merely involve ordered trial-and-error analysis. A few of the time-domain methods are described here.

Evan's root-locus method^(6,7) can be used for the design of compensation functions by postulating a series of trial forms of the proposed compensation functions and plotting the root locus for each form (see Pars. 4-4 and 5-7 for the technique of root-locus construction and the nature of the degree-of-stability criterion). On adjusting the gain to satisfy the degree-of-stability criterion with a specified damping ratio ζ for the dominant pole pair, each trial root locus will produce a specific closed-loop pole-zero configuration. Then by direct inspection of these configurations or by plots of the actual transient responses (through partial-fraction expansion and inverse Laplace transformation), the best compensation form may be selected.

Yeh⁽¹³⁾ has proposed an extension of Evan's method which involves plotting contours of closed-loop pole location for a series of *fixed-gain* values as some parameter of the compensation function is varied. These plots are called

THEORY

gain contours. In addition, for *fixed* values of the compensation-function parameter, contours of closed-loop pole location are plotted as the gain is varied. These plots are called *root contours*. By examining the gain and root contours, the best combination of the compensation-function parameter and the gain can be selected.

Truxal⁽⁶⁾ has developed a pure synthesis procedure based on the desired closed-loop pole-zero configuration (see Par. 7-1). It is assumed that this configuration is characterized by: (1) one pair of conjugate-complex dominant poles, (2) one or more dipoles (a pole and zero very close together on the negative real axis), (3) poles on the negative real axis that are far removed from the dominant pole pairs, and (4) one or more finite zeros⁽⁹⁾. This closed-loop pole-zero configuration will be produced by an open-loop function $[C(s)/E(s)]$ for a unity feedback system] which has all its poles on the negative real axis. If the closed-loop function

$$\frac{C(s)}{R(s)} = \frac{P(s)}{N(s)} \quad (6-36)$$

and, for a unity-feedback system,

$$\frac{C(s)}{E(s)} = \frac{P(s)}{Q(s)} \quad (6-37)$$

then

$$Q(s) = N(s) - P(s) \quad (6-38)$$

where $P(s)$, $N(s)$, and $Q(s)$ are polynomials in s .

The synthesis procedure is then a method for determining the zeros of Eq. (6-38) since these are the poles of $C(s)/E(s)$. Since all the poles of $C(s)/E(s)$ lie on the negative real axis, if the polynomials $N(s)$ and $P(s)$ are plotted on the same coordinate system for $s = -\sigma$ where σ is a real variable, then the intersections of the two curves give the poles of $C(s)/E(s)$. The zeros of $C(s)/E(s)$ are the same as the zeros of $C(s)/R(s)$. Knowing the transfer function of the fixed elements $G_f(s)$, the compensation can be determined from the following equation:

$$G_c(s) = \frac{1}{G_f(s)} \left[\frac{C(s)}{E(s)} \right] \quad (6-39)$$

The cancellation of the function $G_f(s)$ by the compensation $G_c(s)$ should be avoided as much as possible by having some of the poles of $G_f(s)$ occur in the open-loop function $C(s)/E(s)$. This can be accomplished by altering slightly the specified form of the closed-loop response $C(s)/R(s)$ since the performance specifications are rarely rigid. Changes in the parameters of the fixed elements $G_f(s)$ will negate the cancellation called for by Eq. (6-39). Actually, exact cancellation is not necessary since small parameter variations will not alter the closed-loop response appreciably.

6-6 TYPICAL COMPENSATION NETWORKS (2,5,6,14,15,16,17,18,19,20,21)

6-6.1 D-C ELECTRIC

The most common d-c networks are the lag network and the lead network shown in Fig. 6-16.

The *lag-network* transfer function is

$$\frac{E_o(s)}{E_i(s)} = \frac{Ts + 1}{\alpha Ts + 1} \quad (6-40)$$

where

$$T = R_2 C, \text{ and} \quad (6-41)$$

$$\alpha = 1 + \frac{R_1}{R_2} \quad (6-42)$$

The *lead network* transfer function is

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{\alpha} \frac{\alpha Ts + 1}{Ts + 1} \quad (6-43)$$

where

$$T = \left(\frac{R_1 R_2}{R_1 + R_2} \right) C, \text{ and} \quad (6-44)$$

$$\alpha = 1 + \frac{R_1}{R_2} \quad (6-45)$$

COMPENSATION TECHNIQUES

Chestnut and Mayer⁽⁶⁾ present a series of charts of d-c networks containing only resistances and capacitances. These charts cover most of the desirable frequency-response characteristics that are called for in compensation of feedback control systems.

6-6.2 A-C ELECTRIC (2, 5, 6, 14, 15, 16, 17, 18, 19, 20)

In many control system applications, the signals are suppressed-carrier modulated, and the control information modulates a constant-amplitude carrier signal (in practice, 60 or 400 cps). For example, the form of voltage corresponding to the actuating signal may be as follows:

$$V(t) = e(t) \cos \omega_o t$$

where $e(t)$ is the true data signal, and ω_o is the carrier frequency. Networks which are designed to operate on the data of carrier-modulated signals are called a-c or carrier-frequency networks. If it is necessary to compensate a system employing carrier-modulated signals a-c networks are required since d-c networks will not work because they effectively operate on zero-frequency-carrier signals.

There are two questions involved in treating the compensation of carrier-modulated signals:

(1) Analysis: Given a network which operates on a carrier-modulated signal, what is the data-frequency (d-c) equivalent network?

(2) Synthesis: Given a desired data-frequency (d-c) network, what is the equivalent carrier-frequency network?

If $H(j\omega)$ is the frequency response of a carrier network, the frequency response of the data-frequency equivalent is given by

$$H(j\omega_d) = \frac{1}{2} \sqrt{A^2 + B^2} e^{j\psi_d} \quad (6-46)$$

where

$$A = |H_+| \cos(\angle H_+) + |H_-| \cos(\angle H_-) \quad (6-47)$$

$$B = -|H_+| \sin(\angle H_+) + |H_-| \sin(\angle H_-) \quad (6-48)$$

$$\psi_d = \tan^{-1} \frac{B}{A} \quad (6-49)$$

$$H_+ \triangleq H[j(\omega_o + \omega_d)] \quad (6-50)$$

$$H_- \triangleq H[j(\omega_o - \omega_d)] \quad (6-51)$$

ω_o = carrier frequency

ω_d = data frequency

$H(j\omega_d)$ = frequency response of equivalent data-frequency network.

There is no unique answer to the synthesis problem, but a convenient answer is given by the "low-pass to band-pass" transformation. If it is assumed that the magnitude of the carrier-frequency equivalent of a data-frequency network has even symmetry about the

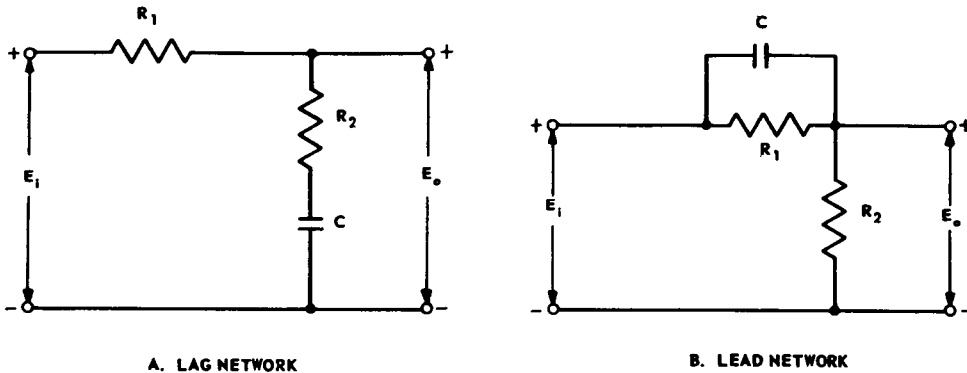


Fig. 6-16 D-C compensation networks.

THEORY

carrier frequency ω_o and that the phase angle has odd symmetry about ω_o , i.e.,

$$|H_+| = |H_-| \quad (6-52)$$

$$\angle H_+ = -\angle H_-, \quad (6-53)$$

then the carrier-frequency equivalent to a given data-frequency network is

$$H(j\omega) = H[j(\omega - \omega_o)] \quad (6-54)$$

Unfortunately, an exact solution of Eq. (6-54) leads to the conclusion that if the data-frequency network H is physically realizable, the carrier-frequency network H is not. However, an approximation to the low-pass to band-pass transformation is possible which does lead to physically realizable networks. If the data frequencies ω_d are small compared to the carrier frequency ω_o , then

$$j(\omega - \omega_o) \cong j \frac{\omega}{2} + \frac{\omega_o^2}{2j\omega} \quad (6-55)$$

As an example, suppose that the frequency response of a data-frequency network is given by

$$H(j\omega_d) = \frac{T_d j\omega_d + 1}{T_d j\omega_d + 1} \quad (6-56)$$

The carrier-frequency equivalent can be found by using Eqs. (6-54) and (6-55). Thus

$$H(j\omega) \cong \frac{T_n \left[j \frac{\omega}{2} + \frac{\omega_o^2}{2j\omega} \right] + 1}{T_d \left[j \frac{\omega}{2} + \frac{\omega_o^2}{2j\omega} \right] + 1} \quad (6-57)$$

There are several ways to realize a carrier-frequency network which is approximately equivalent to a given data-frequency network. The resistance-inductance-capacitance realization starts with the actual data-frequency circuit and replaces the data-frequency circuit elements by their approximate carrier-frequency equivalents as shown in Fig. 6-17. Because of the practical difficulty of realizing parallel inductance-capacitance combinations in the carrier-frequency network, the usual procedure is to realize the data-frequency transfer function by means of a resistance-inductance circuit. Then the carrier equivalent will contain only series inductance-capacitance combinations.

Because lag networks are usually inserted at very low data frequencies, their carrier equivalents are required to be very sharply tuned to the carrier frequency. That is, the carrier equivalent network must be a high-Q circuit. Unfortunately, such high-Q circuits are impractical for servo carrier frequencies (60 and 400 cps) and are very sensitive to

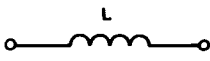
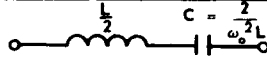
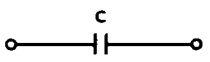
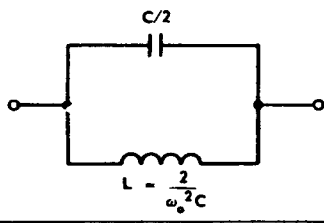


| DATA-FREQUENCY ELEMENT | CARRIER-FREQUENCY EQUIVALENT |
|---|--|
|  $Z = j\omega_d L$ |  |
|  $Y = j\omega_d C$ |  |
|  |  |

Fig. 6-17 Equivalent circuit elements for carrier-frequency networks.

carrier frequency drift. To get around these limitations when lag compensation is desired for a carrier-modulated system, the usual procedure is to demodulate the signal, compensate it with a d-c lag network, and then remodulate.

The realization of carrier-frequency lead networks is not as difficult as the realization of lag networks since they operate at relatively high data frequencies and therefore do not require excessively high-Q circuits. However, carrier lead networks are also sensitive to carrier drift, although some attempts have been made to counteract this effect.^(15,18) If the carrier drift is large (more than 5%), then the scheme of demodulation, compensation, and remodulation should be considered. An effective alternative to this scheme utilizes feedback compensation with a tachometer.

An examination of the magnitude responses of carrier networks shows that they fall into the class of filters called "notch" filters. Methods for realizing notch filters with resistance-capacitance rather than resistance-inductance-capacitance networks are discussed in Refs. (2,5,14). A typical resistance-capacitance notch filter is shown in Fig. 6-18. The frequency response of this carrier network is given by

$$\frac{E_o(j\omega)}{E_i(j\omega)} = H(j\omega) = \frac{\left(j\frac{\omega}{\omega_o}\right)^2 + 1}{\left(j\frac{\omega}{\omega_o}\right)^2 + 4\left(j\frac{\omega}{\omega_o}\right) + 1} \quad (6-58)$$

where

$$\omega_o = \frac{1}{RC} \quad (6-59)$$

The approximate data-frequency equivalent is

$$H(j\omega_d) \cong \frac{j\omega_d}{j\omega_d + 2\omega_o} \quad (6-60)$$

Thus, as far as data frequencies are concerned, the symmetrical parallel $-T$ notch filter behaves as a differentiator for data frequencies up to approximately $0.2\omega_o$.

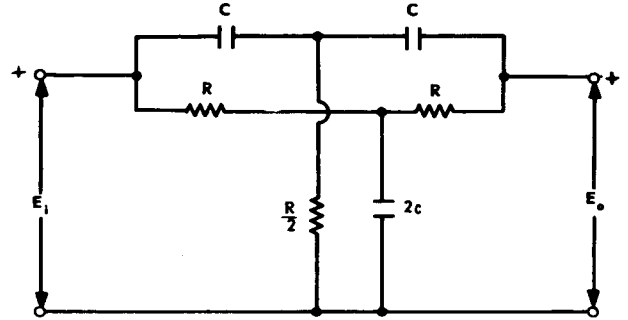


Fig. 6-18 Parallel-T notch filter.

The major difficulty in using resistance-capacitance notch filters is that they must be tuned by successive adjustments of several circuit elements; otherwise, high-precision elements must be used.

6-6.3 MECHANICAL DAMPER

A widely used mechanism having the action of a compensation network is the *inertia damper* shown in Fig. 6-19. The damper, which is connected directly to the servomotor shaft, consists of a thin cylindrical metal shell, a heavy cylindrical metal slug, and a damping fluid. If one neglects motor damping, the block diagram of the inertia damper and servomotor is that shown in Fig. 6-20. In this figure,

$$T = \frac{J_d(J_s + J_m)}{B(J_s + J_m + J_d)} \quad (6-61)$$

$$\alpha = 1 + \frac{J_d}{J_s + J_m} \quad (6-62)$$

T_m = motor torque

T_l = load torque

θ_m = motor shaft position

J_m = motor moment of inertia and reflected load inertia

J_s = shell moment of inertia

J_d = slug moment of inertia

B = fluid damping

The advantages of the inertia damper are:

- (1) simplicity
- (2) no steady power loss

THEORY

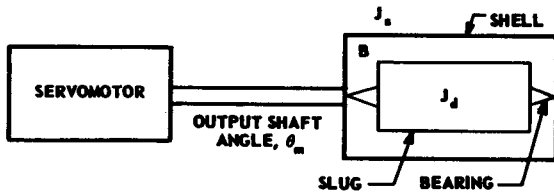


Fig. 6-19 Inertia damper.

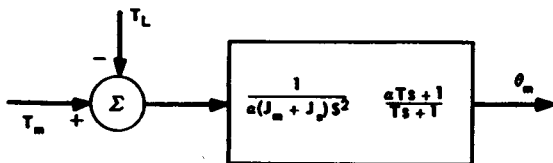


Fig. 6-20 Block diagram of inertia damper (motor damping negligible).

The disadvantages are:

(1) damper must be designed and built for each specific application.

(2) peak acceleration of the damper-motor combination is reduced relative to that of the motor alone because of the added apparent inertia produced at the motor shaft by the damper mechanism.

6-6.4 HYDRAULIC AMPLIFIER (See Par.13-6)

A fairly common means for obtaining *lag-network* action in a hydraulic amplifier is shown in Fig. 6-21. In this figure,

x_i = input displacement of pilot valve

x_o = output motion of power piston

x_s = feedback motion of follow-up sleeve

B = damping of fluid dashpot

K_1, K_2 = spring constants

a, b = lever arms

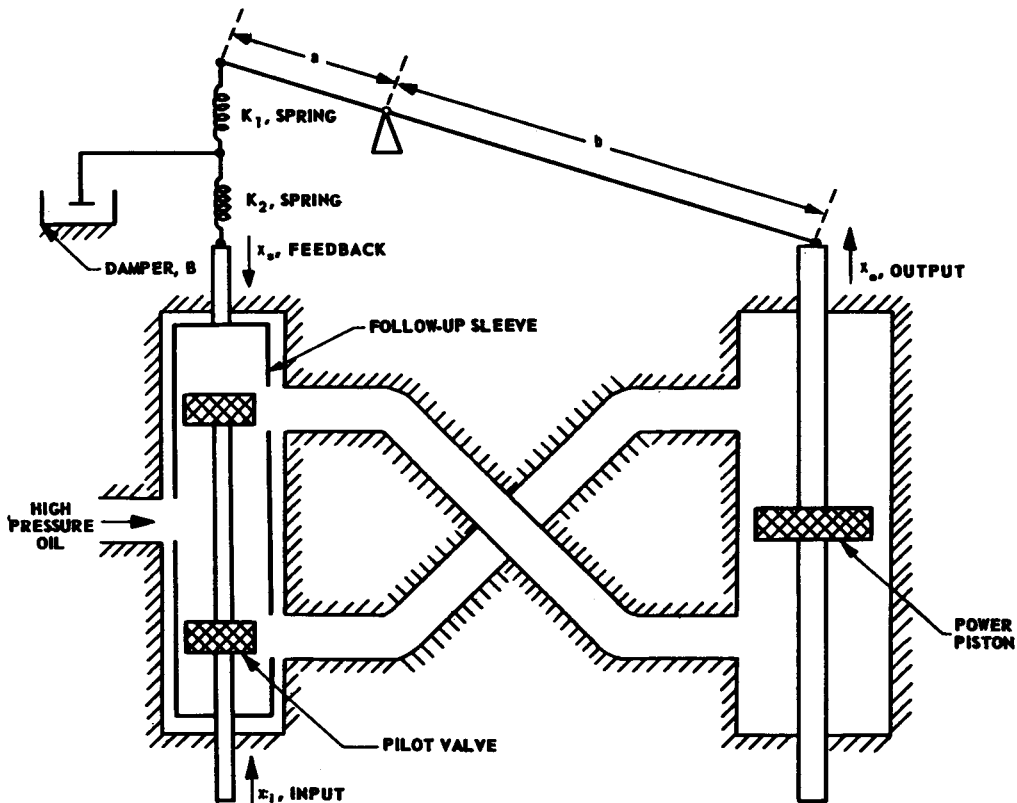


Fig. 6-21 Hydro-mechanical compensation network.

COMPENSATION TECHNIQUES

If the gain of the hydraulic amplifier is very high (greater than 10), the transfer function of the system is

$$\frac{x_o(s)}{x_i(s)} \cong \frac{b}{a} \alpha (\alpha - 1) \left(\frac{1 + Ts}{1 + \alpha Ts} \right) \quad (6-63)$$

where

$$\alpha = 1 + \frac{K_2}{K_1} \quad (6-64)$$

and

$$T = B/(K_2 + K_1) \quad (6-65)$$

6-6.5 PNEUMATIC CONTROLLER⁽²¹⁾

(See Par. 13-7.)

The general schematic of a typical pneumatic controller is shown in Fig. 6-22. In this figure,

- r = motion of set point (reference input)
- c = motion of pen (controlled variable or output)

e = actuating signal

x_f = flapper motion

P_r = nozzle back pressure

P_o = pilot relay output pressure (to diaphragm valve or similar load)

x_{fb} = feedback motion.

If the nozzle-flapper amplifier and the pilot relay are assumed ideal, the block diagram of the controller is as shown in Fig. 6-23, where

K_1 = ratio of proportioning linkage
($0 < K_1 < 1$)

k_{nf} = nozzle-flapper gain

k_{pr} = pilot relay gain

Simple *proportional action* (pure gain) is possible if the feedback function is achieved by means of a spring-loaded bellows as shown in Fig. 6-24. If the ratio of feedback motion x_{fb} to pilot-relay output pressure P_o is denoted

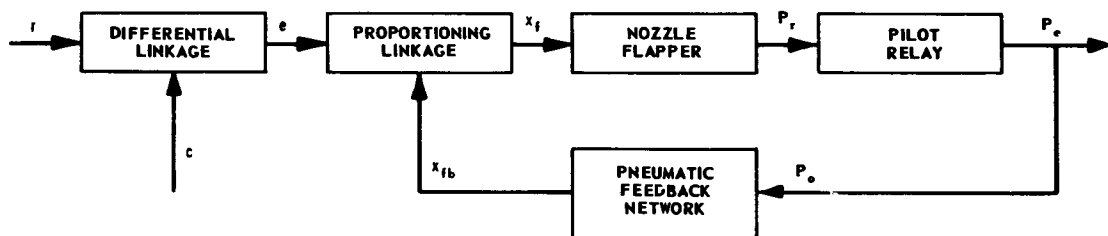


Fig. 6-22 Schematic diagram of a pneumatic controller.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

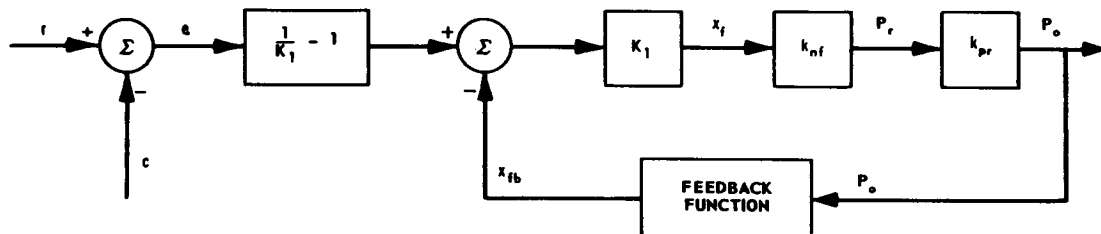


Fig. 6-23 Controller block diagram.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

THEORY

by K_{fb} , the transfer function of the proportional controller is given by

$$\frac{P_o(s)}{e(s)} = \frac{(1 - K_1) k_{nf} k_{pr}}{1 + K_1 K_{fb} k_{nf} k_{pr}} \quad (6-66)$$

If the product $K_1 K_{fb} k_{nf} k_{pr}$ is very high (greater than 10), the proportional controller has the approximate response

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} \quad (6-67)$$

A lag-compensation effect (proportional-plus-integral) can be achieved if the feed-

back function of Fig. 6-22 is obtained by means of the arrangement of Fig. 6-25. The feedback function in this case is given by

$$\frac{x_{fb}(s)}{P_o(s)} = K_{fb} \left(\frac{T_R s}{T_R s + 1} \right) \quad (6-68)$$

where

$$T_R \triangleq R_R C_R$$

R_R = integral resistance

C_R = capacitance of tank

K_{fb} = sensitivity of proportional bellows

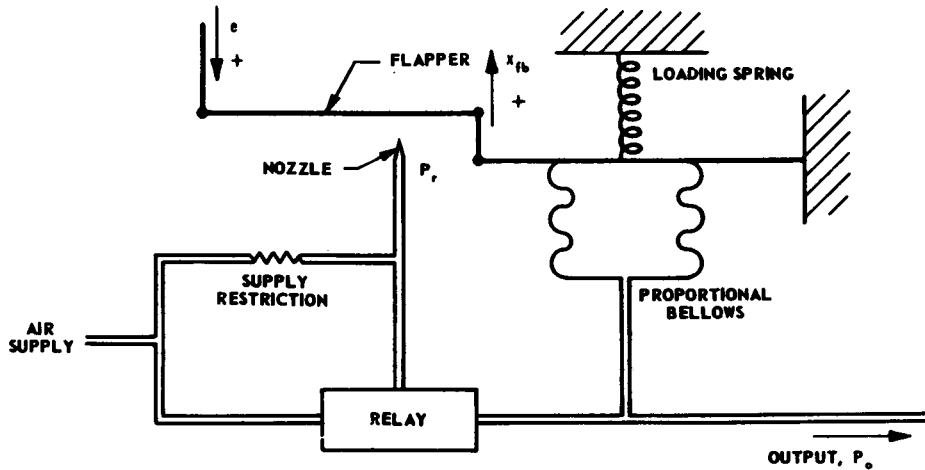


Fig. 6-24 Schematic diagram of a proportional controller.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

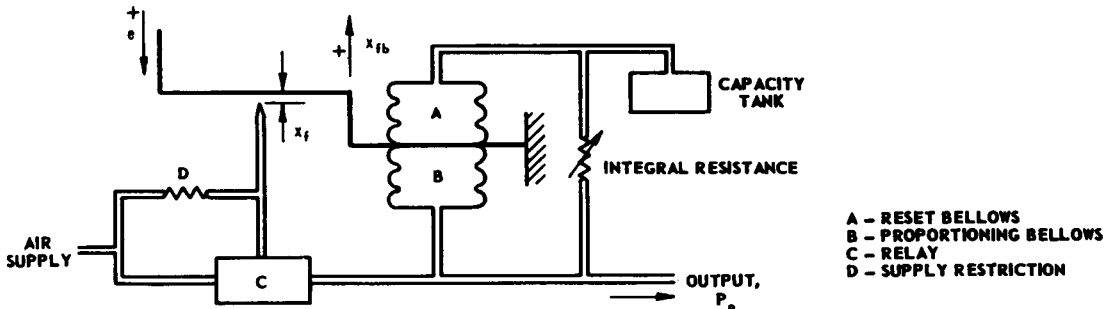


Fig. 6-25 Schematic diagram of a proportional plus integral controller.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

COMPENSATION TECHNIQUES

The transfer function of the controller then becomes

$$\frac{P_o(s)}{e(s)} = (1 - K_1) k_{nf} k_{pr} \left(\frac{T_R s + 1}{\alpha_R T_R s + 1} \right) \quad (6-69)$$

where

$$\alpha_R = 1 + K_1 k_{nf} k_{pr} K_{fb} \quad (6-70)$$

If the product $K_1 k_{nf} k_{pr} K_{fb}$ is very high (greater than 50), the response of the proportional-plus-integral controller is approximately

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} \left(1 + \frac{1}{T_R s} \right) \quad (6-71)$$

The form of the right side of this equation explains the name — “proportional-plus-integral” controller.

A *lead-compensation effect* (proportional-plus-derivative) can be achieved if the feedback function of Fig. 6-22 is obtained by means of the arrangement of Fig. 6-26. The feedback function in this case is given by

$$\frac{x_{fb}(s)}{P_o(s)} = K_{fb} \left(\frac{\frac{T_d}{b} s + 1}{T_d s + 1} \right) \quad (6-72)$$

where

$$T_d \triangleq R_d C_d \quad (6-73)$$

$$b \triangleq 1 + \frac{A_d}{A_p} \quad (6-74)$$

R_d = derivative resistance

C_d = capacitance of tank

A_d = area of derivative bellows

A_p = area of proportional bellows

K_{fb} = sensitivity of proportional bellows

The transfer function of this controller is

$$\frac{P_o(s)}{e(s)} = \frac{(1 - K_1) k_{nf} k_{pr}}{1 + K_1 k_{nf} k_{pr} K_{fb}} \left(\frac{T_d s + 1}{\frac{T_d}{\alpha_d} s + 1} \right) \quad (6-75)$$

where

$$\alpha_d = \frac{1 + K_1 k_{nf} k_{pr} K_{fb}}{1 + \frac{1}{b} (K_1 k_{nf} k_{pr} K_{fb})} \quad (6-76)$$

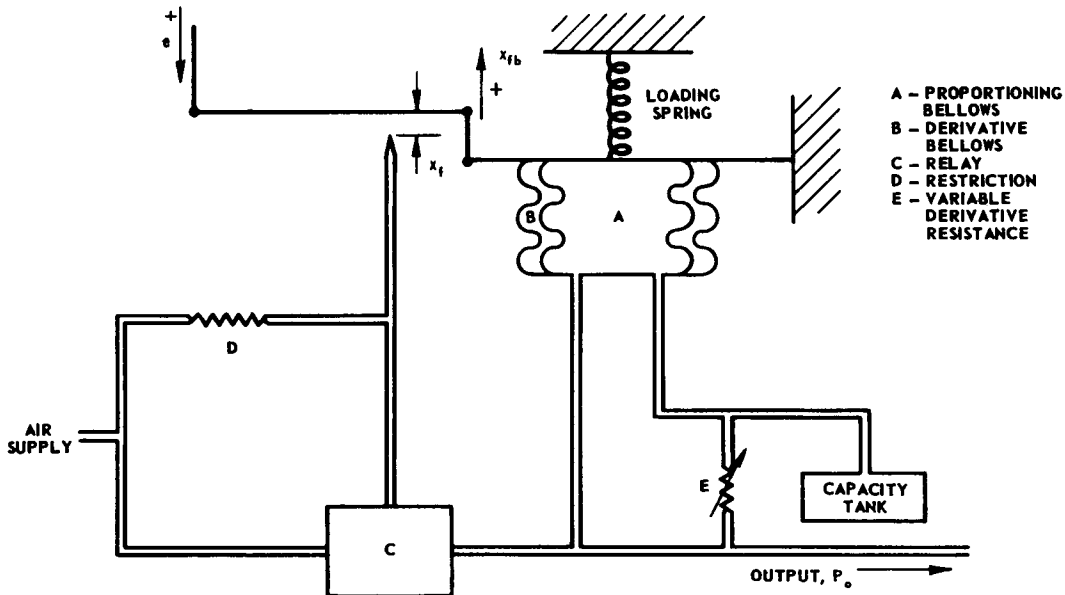


Fig. 6-26 Schematic diagram of proportional plus derivative controller.

Adapted by permission from *Instruments*, Volume 26, No. 6, June, 1953, from article entitled 'Dynamic Behavior of Pneumatic Devices', by L. A. Gould and P. E. Smith, Jr.

THEORY

If the product $\frac{1}{b} (K_1 k_n k_{pr} K_{fb})$ is very high (greater than 10), the response of the proportional-plus-derivative controller is approximately

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} \left(\frac{T_d s + 1}{\frac{T_d}{b} s + 1} \right) \quad (6-77)$$

If the ratio of the areas (A_d/A_p) is very high

(greater than 50), then $b \gg 1$, and the response is given approximately by

$$\frac{P_o(s)}{e(s)} \cong \frac{(1 - K_1)}{K_1 K_{fb}} (1 + T_d s) \quad (6-78)$$

The form of the right side of this equation explains the term — “proportional-plus-derivative” controller.

BIBLIOGRAPHY

- 1 Edited by I. A. Greenwood, Jr., J. V. Holdam, Jr., and D. MacRay, Jr., *Electronic Instruments*, MIT Radiation Laboratory Series, Vol. 21, pp. #319-344, McGraw-Hill Book Company, Inc., New York, N. Y., 1948.
- 2 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms*, MIT Radiation Laboratory Series, Vol. 25, pp. #114-130, 196-211, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.
- 3 G. S. Brown and D. P. Campbell, *Principles of Servomechanism*, pp. #195-235, 262-292, John Wiley & Sons, Inc., New York, N. Y., 1948.
- 4 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. 1, pp. #245-290, 327-357, John Wiley & Sons, Inc., New York, N. Y., 1951.
- 5 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. 2, pp. #119-149, 187-213, John Wiley & Sons, Inc., New York, N. Y., 1955.
- 6 J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. #250-266, 278-409, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.
- 7 W. R. Evans, “Control System Synthesis by Root-Locus Method”, *Trans. AIEE*, Vol. 69, Part I, pp. #66-69, 1950.
- 8 P. Travers, “A Note on the Design of Conditionally Stable Feedback Systems”, *Trans. AIEE*, Vol. 70, Part I, pp. #626-630, 1951.
- 9 M. R. Aaron, “Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed Loop Function”, *Trans. AIEE*, Vol. 70, Part II, pp. #1439-1445, 1951.
- 10 D. W. Russel and C. H. Weaver, “Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location”, *Trans. AIEE*, Vol. 71, Part II, pp. #95-104, 1952.
- 11 J. M. Smith, “Stabilization Templates for Servomechanisms”, *Trans. AIEE*, Vol. 71, Part II, pp. #220-224, 1952.
- 12 J. R. Koenig, “A Relative Damping Criterion for Linear Systems”, *Trans. AIEE*, Vol. 72, Part II, pp. #291-294, 1953.
- 13 V. C. M. Yeh, “Synthesis of Feedback Control Systems by Gain-Contour and Root-Contour Methods”, *Trans. AIEE*, Vol. 75, Part II, pp. #85-96, 1956.

COMPENSATION TECHNIQUES

- 14 A. Sobczyk, "Stabilization of Carrier Frequency Servomechanisms", *J. Franklin Inst.*, Vol. 246, pp. #21-43, 95-121, 187-213, July-August-September, 1948.
- 15 A. P. Notthoff, Jr., "Phase Lead for A-C Servo Systems with Compensation for Carrier Frequency Changes", *Trans. AIEE*, Vol. 69, Part I, pp. #285-291, 1950.
- 16 D. McDonald, "Improvements in the Characteristics of A-C Lead Networks for Servomechanisms", *Trans. AIEE*, Vol. 69, Part I, pp. #293-300, 1950.
- 17 R. L. Cosgriff, "Integral Controller for Use in Carrier-Type Servomechanisms", *Trans. AIEE*, Vol. 69, Part II, pp. #1379-1383, 1950.
- 18 G. M. Attura, "Effects of Carrier Shifts on Derivative Networks for A-C Servomechanisms", *Trans. AIEE*, Vol. 70, Part II, pp. #612-618, 1951.
- 19 G. A. Bjornson, "Graphical Synthesis by Graphical Methods for A-C Servomechanisms", *Trans. AIEE*, Vol. 70, Part I, pp. #619-625, 1951.
- 20 S. S. L. Chang, "Transient Analysis of A-C Servomechanisms", *Trans. AIEE*, Vol. 74, Part II, pp. #30-37, 1955.
- 21 L. A. Gould and P. E. Smith, Jr., "Dynamic Behavior of Pneumatic Devices", *Instruments*, Vol. 26, No. 6 and 7, pp. #886-891, 1026-1033, 1953.

CHAPTER 7

PERFORMANCE EVALUATION*

7-1 RELATIONS BETWEEN FREQUENCY RESPONSE
AND TRANSIENT RESPONSE

7-1.1 GENERAL

As stated in Par. 6-5, the relationship between transient response and frequency response is somewhat tenuous. Consequently, it is often necessary to have explicit knowledge of the response in both the time and frequency domains. This section presents some of the important approximations that enable the designer to translate between the time- and frequency-domain descriptions of performance. By the use of these approximations, a quick evaluation of performance can be made.

7-1.2 CLOSED-LOOP FREQUENCY RESPONSE
FROM CLOSED-LOOP TRANSIENT
RESPONSE

If the closed-loop *transient response* of a system is known from experimental test data, there are several methods^(21,22,27,28,29,30) available for determining the frequency response.

If the step response of the system is non-oscillatory (i.e., has no overshoot), the transient component of the response can be obtained by subtracting the step response from the final value of the output, i.e.,

$$c_t(t) = c(\infty) - c(t) \quad (7-1)$$

where $c(\infty)$ is the final value of the output, $c(t)$ is the step response, and $c_t(t)$ is the transient component of the response. The logarithm of $c_t(t)$ is plotted against time on semi-log paper. If the response is dominated by an exponential component, the resultant curve plotted on semi-log paper eventually

approaches a straight line whose slope corresponds to the magnitude of the dominant time constant. That is, if the dominant transient component is

$$c_{t_1}(t) = Ae^{-t/T}, \quad (7-2)$$

then

$$\log_e c_{t_1}(t) = \log_e A - t/T \quad (7-3)$$

An extrapolation of the straight-line asymptote of $\log_e c_{t_1}(t)$ back to zero time yields the logarithm of the amplitude A of the dominant transient component. Thus, the dominant transient component is completely determined and can be subtracted from $c_t(t)$. A plot of the logarithm of the difference $[c_t(t) - c_{t_1}(t)]$ versus time t produces a curve which approaches a straight-line asymptote whose slope corresponds to the time constant of the exponential component having the next smaller time constant. Extrapolating this curve back to zero time yields the logarithm of the amplitude of the secondary component, $c_{t_2}(t)$. Next, the function $[c_t(t) - c_{t_1}(t) - c_{t_2}(t)]$ is determined, and the process can be repeated until the limit of measurement accuracy is reached.

Thal-Larsen⁽²¹⁾ gives a method for determining approximate transfer functions based on the approximation of a nonoscillatory step response by the transfer function

$$W(s) = \frac{C(s)}{R(s)} = \frac{e^{-t_d s}}{(s + 1)(T_2 s + 1)(T_3 s + 1)} \quad (7-4)$$

*By L. A. Gould

THEORY

where T_2 and T_3 are dimensionless time constants and t_0 is the dead time.

By choosing the 10%, 40%, and 80% times in the step responses of this system for the various combinations of its parameters, a set of dimensionless curves have been constructed. In using the curves (Figs. 7-1 through 7-5), the three points corresponding to the 10%, 40%, and 80% response levels of the measured response are determined, and the times corresponding to these points are designated t_1 , t_2 , and t_3 , respectively. The values of the dimensionless ratio $(t_3 - t_1)/(t_2 - t_1)$ and the time $(t_3 - t_1)$ together with curves of Figs. 7-4 and 7-5 will enable the designer to determine a set of roots that corresponds to a transient passing through the three selected points. If a dead time t_0 is present, the ratio $(t_3 - t_1)/(t_2 - t_0)$ will enable the designer to select roots that reproduce the first 10% of the transient.

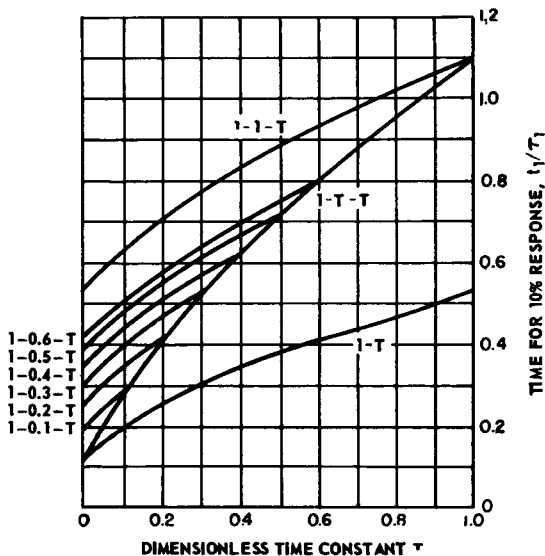


Fig. 7-1 Normalized curves yielding time for 10-percent transient response corresponding to combinations of various time constants

By permission from *Transactions of the AIEE*, Volume 74 Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

Example. Let

$$t_1 = 0.97 \text{ sec}$$

$$t_2 = 2.14 \text{ sec}$$

$$t_3 = 4.47 \text{ sec}$$

Then,

$$\frac{t_3 - t_1}{t_2 - t_1} = 3.00$$

(a) Entering Fig. 7-4 at this value (i.e., 3.00), several curves are crossed allowing the choice of various combinations of the dimensionless or relative time constant T . Choosing three of these combinations:

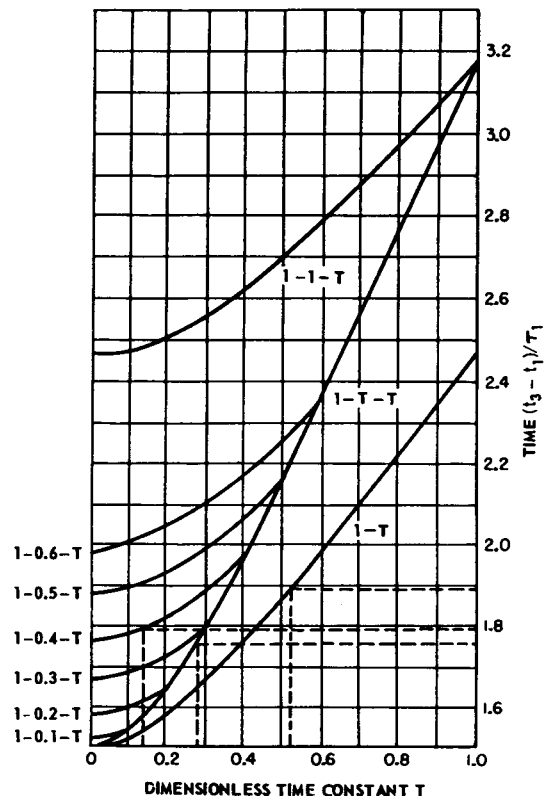


Fig. 7-2 Normalized curves yielding time for 40-percent transient response corresponding to combinations of various time constants.

By permission from *Transactions of the AIEE*, Volume 74, Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

PERFORMANCE EVALUATION

(1) Curve 1 — $T = T$: 1, 0.275, 0.275

(2) Curve 1 — 0.4 — T : 1, 0.4, 0.135

(3) Curve 1 — T : 1, 0.520

(b) Entering Fig. 7-2 with the dimensionless time constants found above, the dimensionless time $(t_3 - t_1)/\tau_1$ is determined:

(1) Curve 1 — $T = T$ for $T = 0.275$: 1.755

(2) Curve 1 — 0.4 — T for $T = 0.135$: 1.790

(3) Curve 1 — T for $T = 0.520$: 1.890

(c) The time $(t_3 - t_1)$ from the actual transient divided by the dimensionless time $(t_3 - t_1)/\tau_1$ yields the conversion factor τ_1 by which the relative time constants found in the first step must be multiplied to obtain the

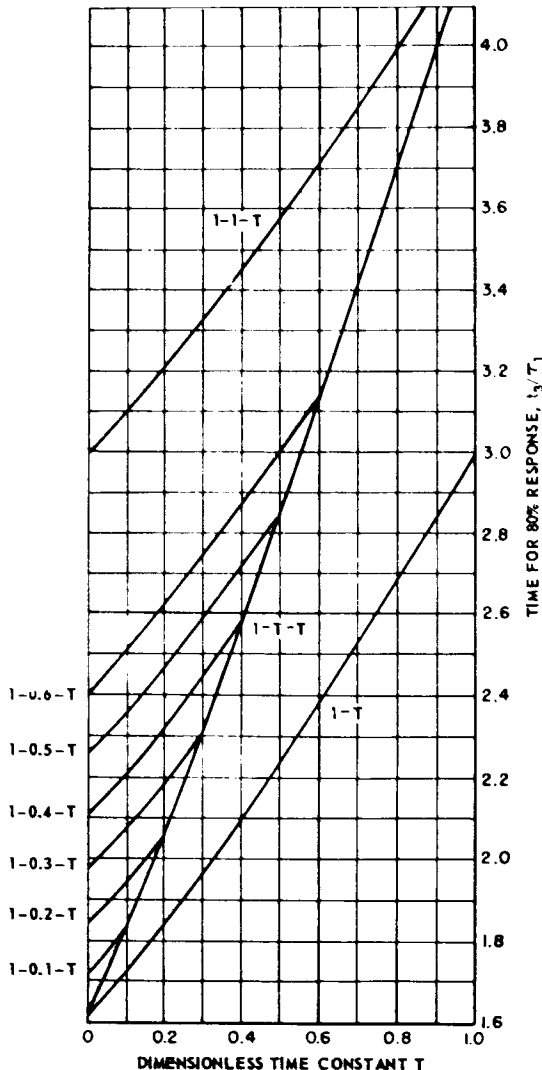


Fig. 7-3 Normalized curves yielding time for 80-percent transient response corresponding to combinations of various time constants.

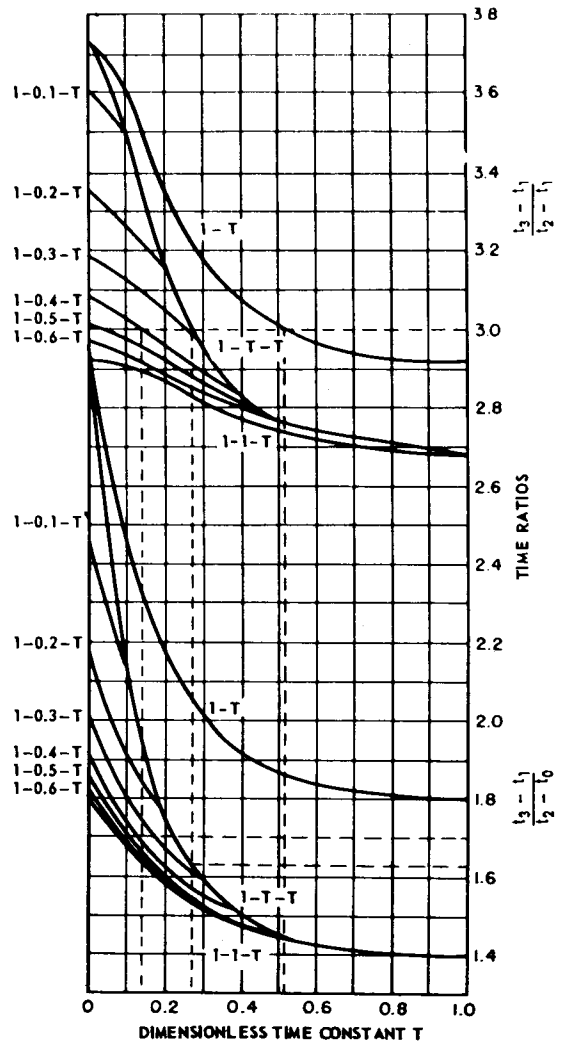


Fig. 7-4 Normalized curves yielding time-interval ratios of the transient response corresponding to combinations of various time constants.

By permission from *Transactions of the AIEE*, Volume 74, Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

By permission from *Transactions of the AIEE*, Volume 74, Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

THEORY

actual time constants. Note that Eq. (7-4) represents a normalized transfer function with dimensionless time constants 1, T_2 , and T_3 . The time constants for the original function before normalization are τ_1 , $\tau_1 T_2$, and $\tau_1 T_3$. For the example $(t_3 - t_1) = 3.50$ sec, the three combinations which fit the original curve are:

- (1) $\tau_1 = 1.995$ sec; $\tau_1 T_2 = \tau_1 T_3 = 0.549$ sec
- (2) $\tau_1 = 1.955$ sec; $\tau_1 T_2 = 0.782$ sec;
 $\tau_1 T_3 = 0.264$ sec
- (3) $\tau_1 = 1.850$ sec; $\tau_1 T_2 = 0.963$ sec;
 $\tau_1 T_3 = 0$

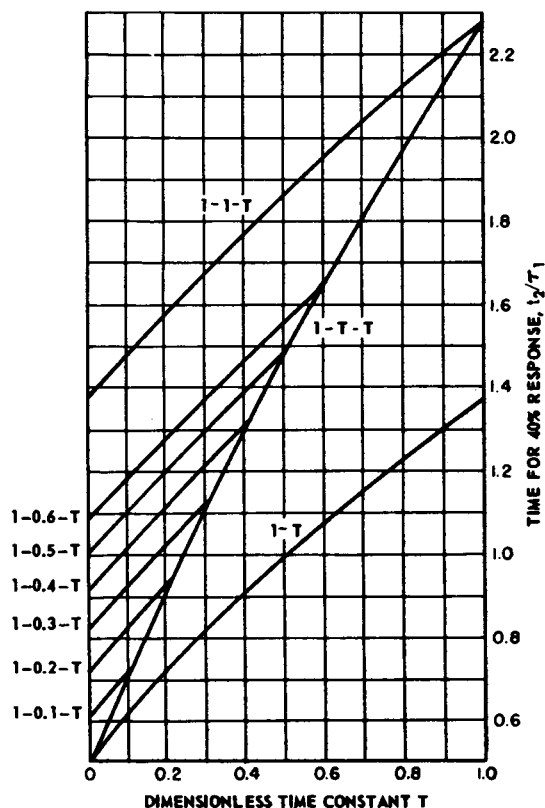


Fig. 7-5 Normalized curves yielding the time interval between 10- and 80-percent response of the transient corresponding to combinations of various time constants.

by permission from *Transactions of the AIEE*, Volume 74, Part II, 1955, from article entitled 'Frequency Response from Experimental Nonoscillatory Transient-Response Data', by H. Thal-Larsen.

(d) To check for the necessity of a dead-time factor, enter Fig. 7-5 with the dimensionless time constants T from the second step to determine t_2/τ_1 .

- (1) Curve $1 - T - T$ for $T = 0.275$; 1.075
- (2) Curve $1 - 0.4 - T$ for $T = 0.135$; 1.055
- (3) Curve $1 - T$ for $T = 0.520$; 1.018

(e) The conversion factor τ_1 found in the third step, together with the results of the fourth step, permit the calculation of the actual time t_2 if no dead time is present. Thus, for the three combinations considered, there results

- (1) $t_2 = 2.142$ sec
- (2) $t_2 = 2.060$ sec
- (3) $t_2 = 1.885$ sec

(f) The times found in the fifth step when subtracted from the measured time t_2 yield the dead time t_0 . The actual measured time $t_2 = 2.14$ sec. Therefore,

- (1) $t_0 = 2.14 - 2.142 = 0$
- (2) $t_0 = 2.14 - 2.060 = 0.080$ sec
- (3) $t_0 = 2.14 - 1.885 = 0.255$ sec

(g) By substituting the appropriate values from steps (c) and (f) into Eq. (7-4), we find that the three transfer functions which approximate the response in the 10% to 80% interval are

- (1) $W(s) \cong \frac{1}{(1.995s+1)(0.549s+1)^2}$
- (2) $W(s) \cong \frac{e^{-0.080s}}{(1.955s+1)(0.782s+1)(0.264s+1)}$
- (3) $W(s) \cong \frac{e^{-0.255s}}{(1.850s+1)(0.963s+1)}$

Chestnut and Mayer⁽²⁷⁾ give graphical methods that are useful for determining frequency response from transient response in

any case (oscillatory or nonoscillatory). To find the frequency response associated with the *step response* of a system, the time axis of the response is divided into equal intervals. Then a "staircase" approximation (see Fig. 7-6) is made to the step response with each step occurring at the middle of a given time interval. If t_n is the middle of the n th time interval and ΔC_n is the change in the response occurring at t_n , then the frequency response is given by

$$W(j\omega) = \sum_{n=1}^{\infty} \Delta C_n e^{j\omega t_n} \quad (7-5)$$

This equation can be evaluated graphically at each frequency by a "vector" summation.

To find the frequency response corresponding to the *impulse response* of a system (the impulse being approximated experimentally by a short finite pulse), the time scale of the

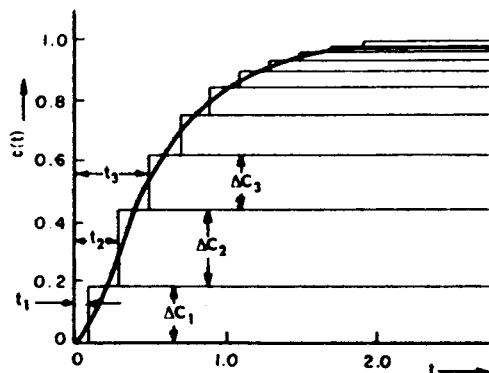


Fig. 7-6 Rectangular approximation to step response.

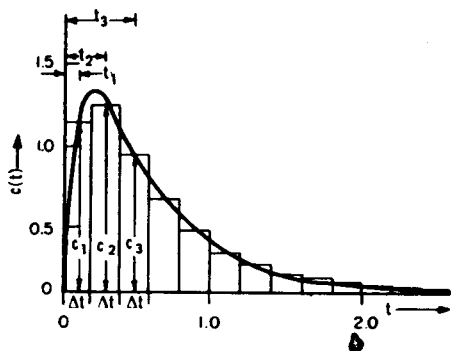


Fig. 7-7 Rectangular approximation to impulse response.

rectangular

impulse response is divided into equal intervals. Then a rectangular-pulse approximation is made to the impulse response (Fig. 7-7). If t_n is the center of the n th time interval, c_n — the value of the impulse response at t_n , and Δt — the length of the time interval, then the frequency response is given by

$$W(j\omega) = \sum_{n=1}^{\infty} c_n \Delta t e^{j\omega t_n} \quad (7-6)$$

This equation can be evaluated graphically at each frequency by a "vector" summation.

Seamans et al.^(28,29) use a triangular method to approximate a given time function $c(t)$. This is equivalent to approximating the time function by straight-line segments and then decomposing the straight-line approximation into a series of isosceles triangles (Fig. 7-8). Once the transform of a single triangular pulse is known, the frequency function $C(j\omega)$ corresponding to $c(t)$ is found from

$$C(j\omega) = e^{j\omega t_0} \left[\frac{\sin \frac{\omega \Delta t}{2}}{\frac{\omega \Delta t}{2}} \right]^2 \Delta t \left[\sum_{n=1}^{\infty} E_n e^{j\omega \Delta t n} \right] \quad (7-7)$$

where t_0 represents the time at the start of the first pulse, Δt — the time interval between pulses, and E_n — the amplitude of the n th pulse.

Guillemin^(28,30) suggests that the time function be approximated by a sequence of rational polynomials in t (straight lines, parabolas,

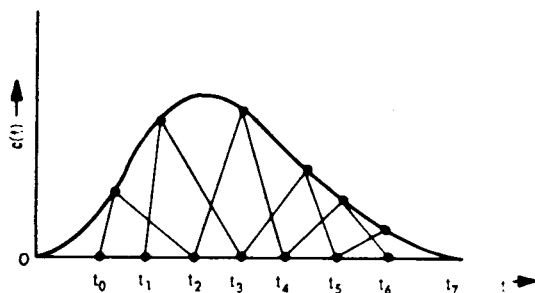


Fig. 7-8 Triangular approximation to time function.

THEORY

cubics, etc.). The approximate function $c^*(t)$ is differentiated enough times (n times) to make $\frac{d^n}{dt^n}c^*$ a sequence of impulses. Actually, the original function may be differentiated before approximating by polynomials so that lower-order polynomials can be used. The final impulse function is then transformed, yielding

$$C(j\omega) \cong \sum_{k=1}^m \frac{a_k e^{-j\omega t_k}}{(j\omega)^n} \quad (7-8)$$

where a_k is the magnitude of the k th impulse, t_k — the time of occurrence of the k th impulse, and n — the total number of times the original function has been differentiated.

If *rational approximations* are sought for an experimentally derived frequency function $W(j\omega)$, advantage can be taken of the fact that the plot of $10 \log_{10} |W(j\omega)|$ vs $\log \omega$ is easily representable by straight-line asymptotes having slopes of $\pm 10n$ dg/dec ($n = 0, 1, 2, \dots$). By combining the straight-line approximation of the magnitude function with the first- and second-order response curves given in Par. 5-3 (Figs. 5-7 through 5-10), curve fitting is possible. The easiest procedure is to use the magnitude curves to get a rough approximation and then to refine the approximation with the phase curves.

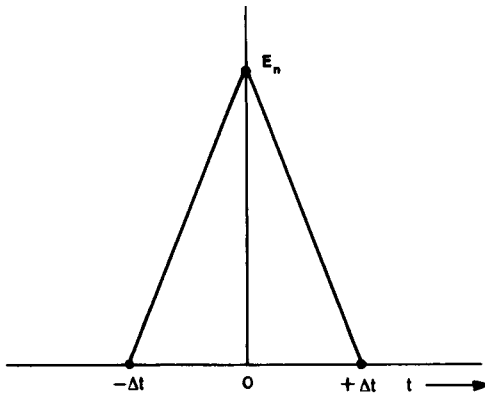


Fig. 7-9 Triangle function.

Linville^(15,28) has proposed a method for improving the foregoing approximation procedure. In this method, an investigation is made of the effect of varying the position of the approximate poles and zeros on the difference between the actual function and the first approximation obtained from fitting the asymptotes and their corresponding correction curves. For example, if

$$F(\omega) = 10 \log_{10} |G(j\omega)| \quad (7-9)$$

and

$$G(s) = \frac{s^2 - 2\sigma_1 s + \sigma_1^2 + \omega_1^2}{s^2 - 2\sigma_2 s + \sigma_2^2 + \omega_2^2} \quad (7-10)$$

then the change in $F(\omega)$ resulting from small changes of the poles ($+\sigma_2 \pm j\omega_2$) and the zeros ($+\sigma_1 \pm j\omega_1$) is given by

$$\Delta F(\omega) = \frac{\partial F}{\partial \sigma_1} \Delta \sigma_1 + \frac{\partial F}{\partial \omega_1} \Delta \omega_1 + \frac{\partial F}{\partial \sigma_2} \Delta \sigma_2 + \frac{\partial F}{\partial \omega_2} \Delta \omega_2 \quad (7-11)$$

The steps in the approximation procedure are as follows:

(a) A plot is made of the difference between the actual $F(\omega)$ and the first approximation in a frequency region where the first approximation is to be improved by changing the position of approximate poles and/or zeros which occur in this region.

(b) The variation of the pertinent partial derivatives of $F(\omega)$ with frequency ω is determined in the vicinity of the approximate poles and/or zeros.

(c) The pertinent partial-derivative curves are used to approximate $\Delta F(\omega)$ in the frequency region of interest. From this approximation, the necessary changes in the positions of the approximate poles and/or zeros are determined.

The curves of Figs. 7-10A through 7-10N can be used to evaluate the necessary partial-derivative curves. Note that the phase corrections can be determined by using the same procedure.

PERFORMANCE EVALUATION

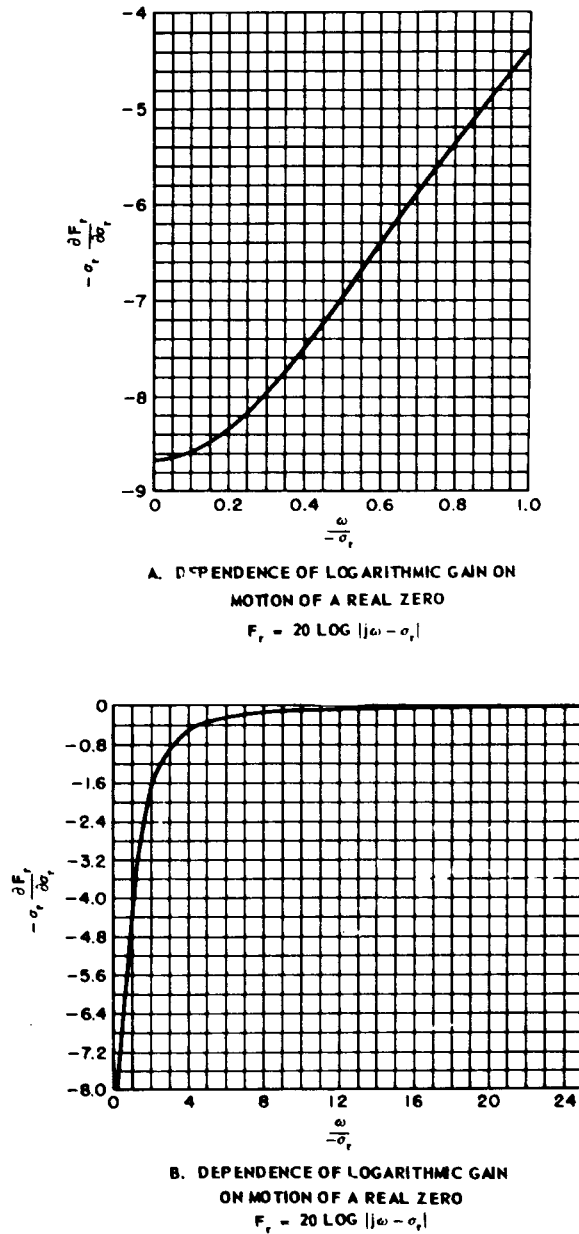
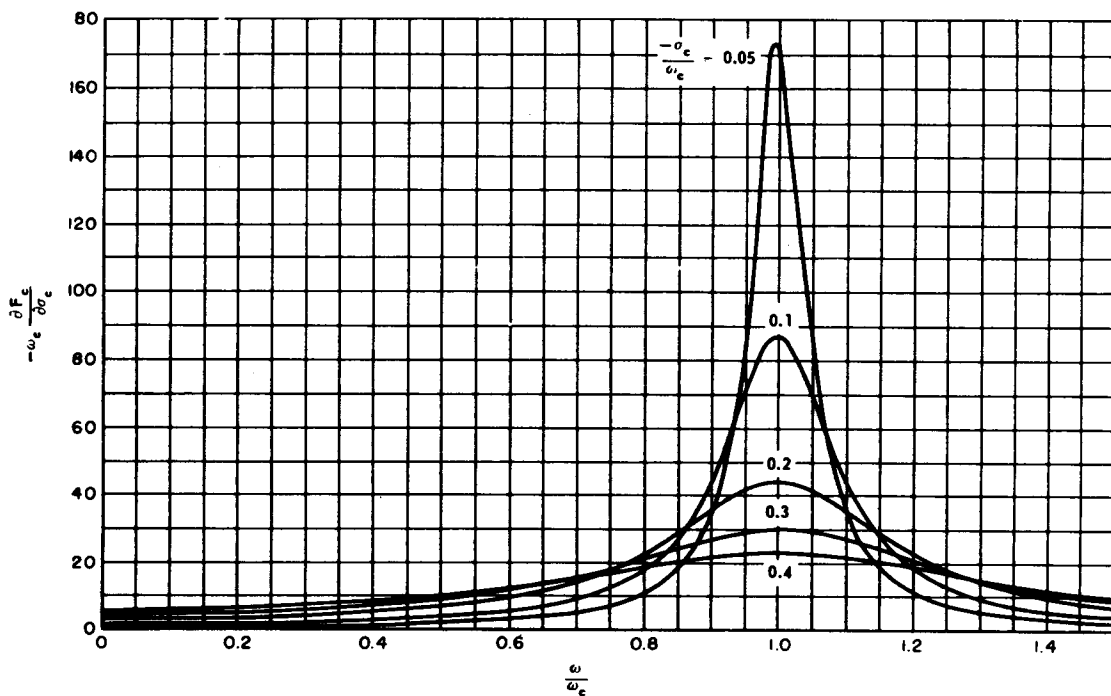


Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 1 of 10)

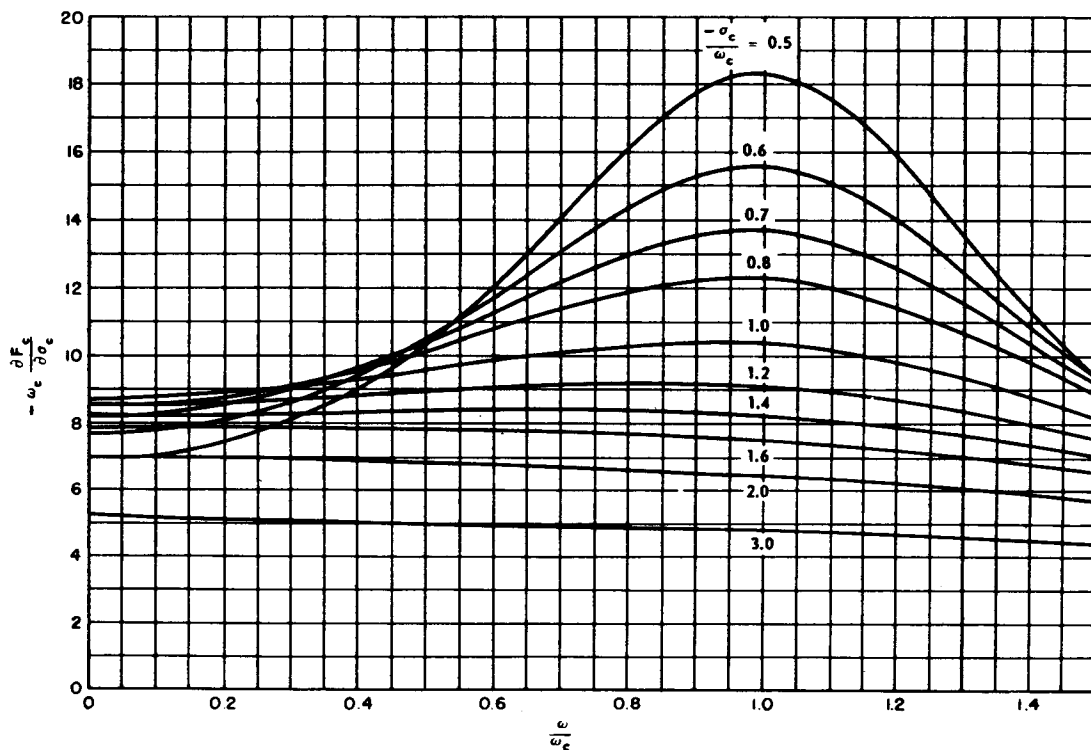
From "The Selection of Network Functions to Approximate Prescribed Frequency Characteristics", by J. G. Linvill, Research Laboratory of Electronics, Technical Report No. 146, March 14, 1950, Massachusetts Institute of Technology.

THEORY



C. DEPENDENCE OF GAIN ON REAL PART OF ZERO (OR POLE)

$$F_c = 20 \text{ LOG } |(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2|$$

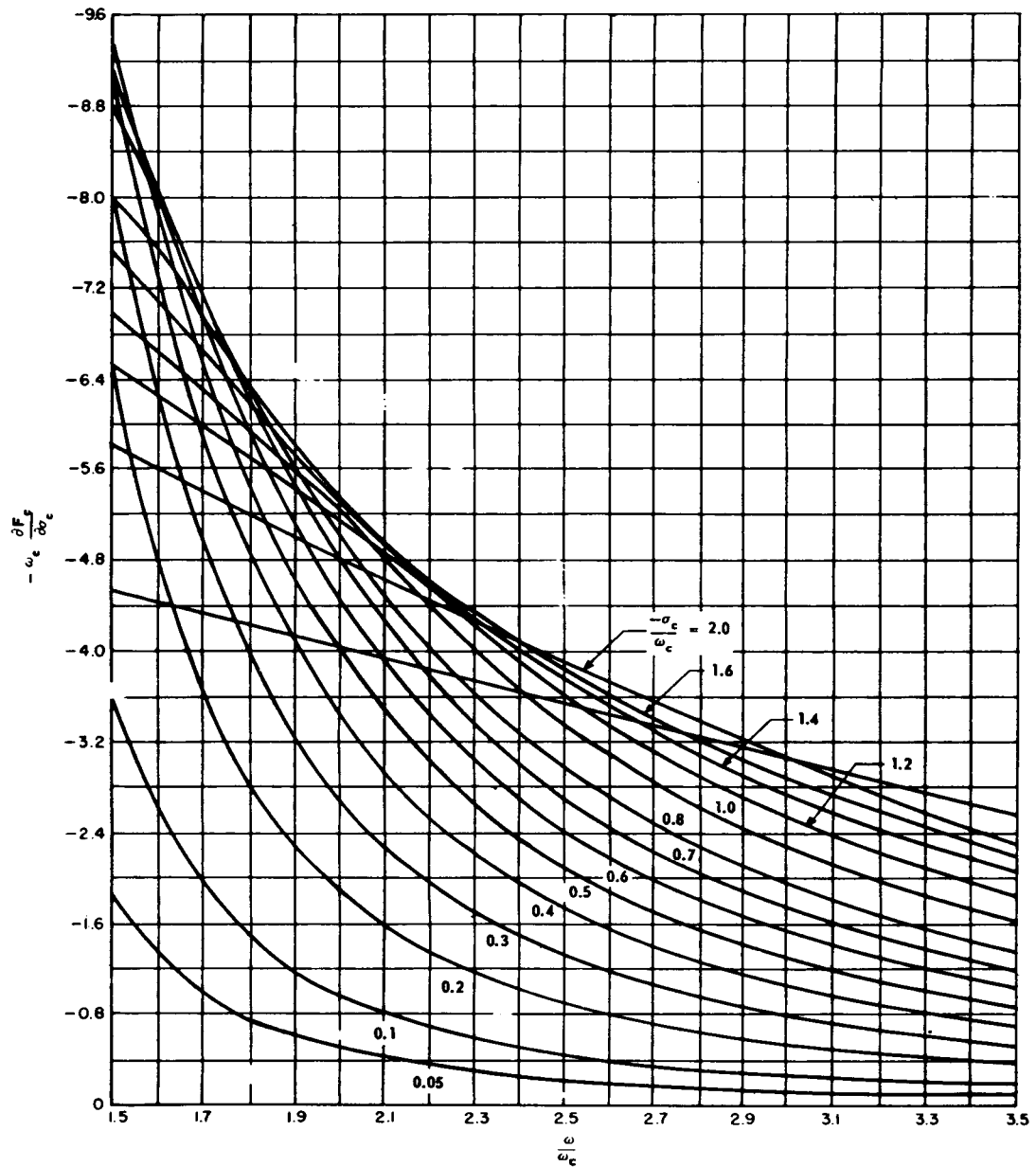


D. DEPENDENCE OF GAIN ON REAL PART OF ZERO (OR POLE)

$$F = 20 \text{ LOG } |(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2|$$

Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 2 of 10)

PERFORMANCE EVALUATION

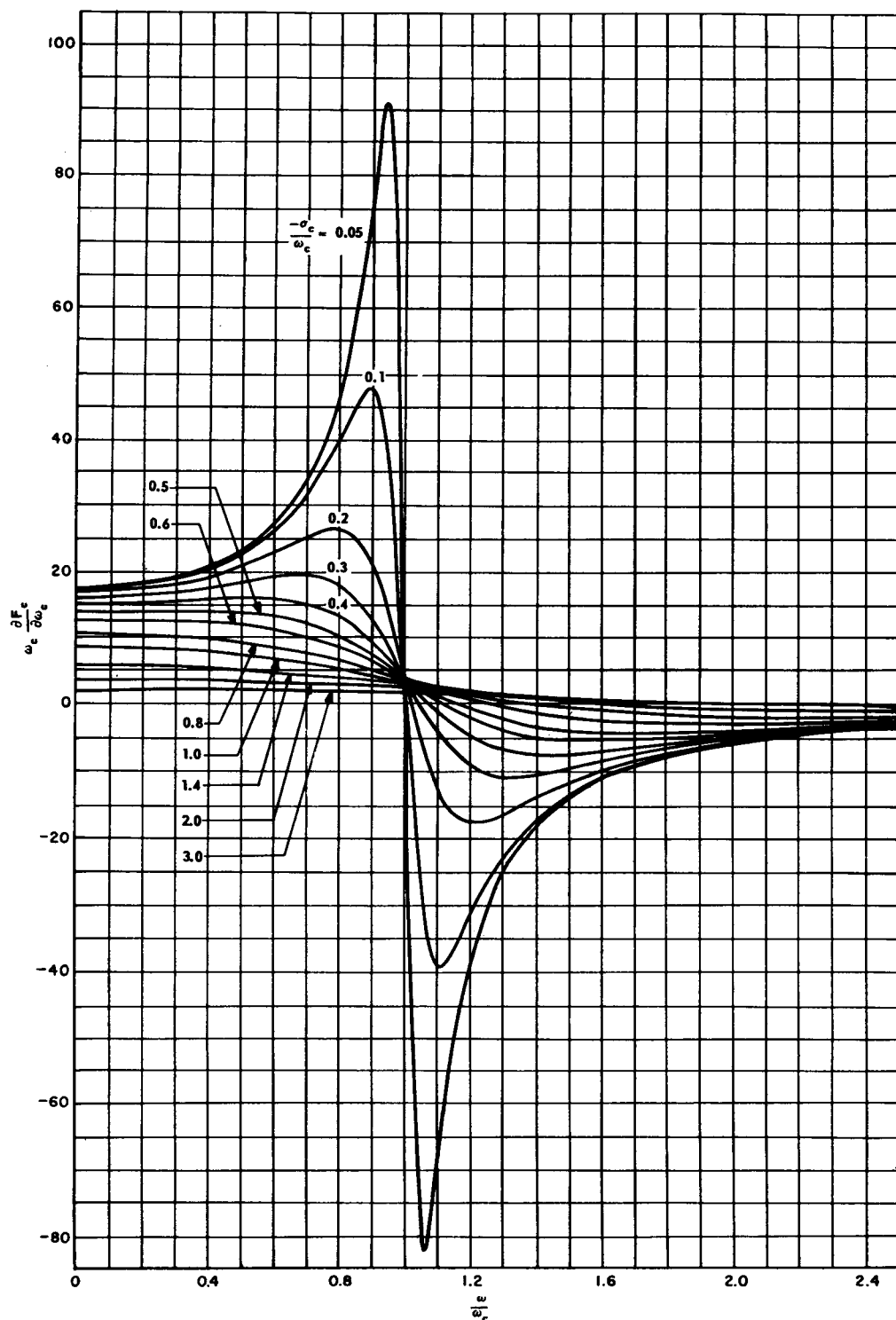


E. DEPENDENCE OF GAIN ON REAL PART OF ZERO (OR POLE)

$$F_c = 20 \text{ LOG } [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$$

Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 3 of 10)

THEORY

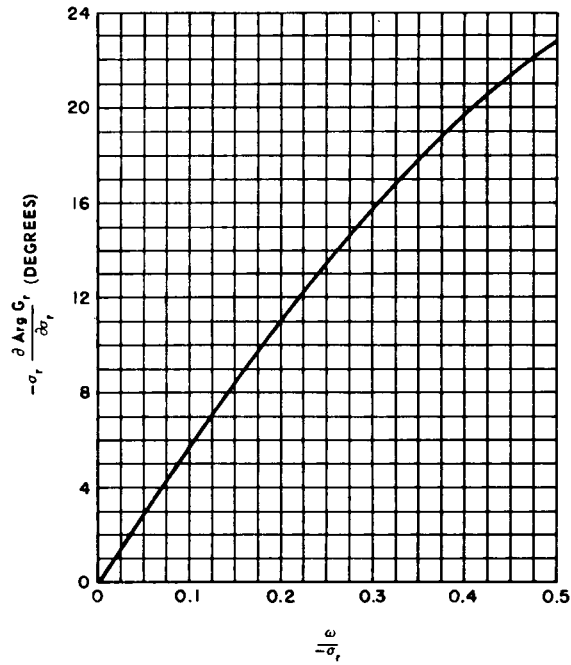


F. DEPENDENCE OF GAIN ON IMAGINARY PART OF ZERO (OR POLE)

$$F_c = 20 \text{ LOG } [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$$

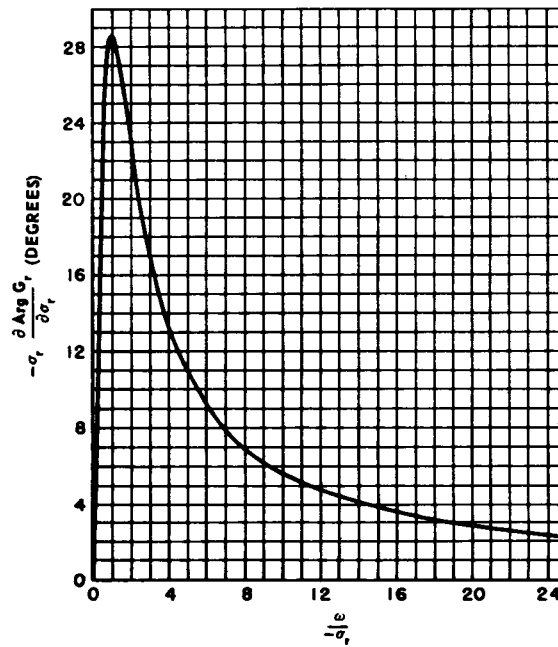
Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 4 of 10)

PERFORMANCE EVALUATION



G. DEPENDENCE OF PHASE ON REAL ZERO (OR POLE)

$$G_r = (j\omega - \sigma_r)$$

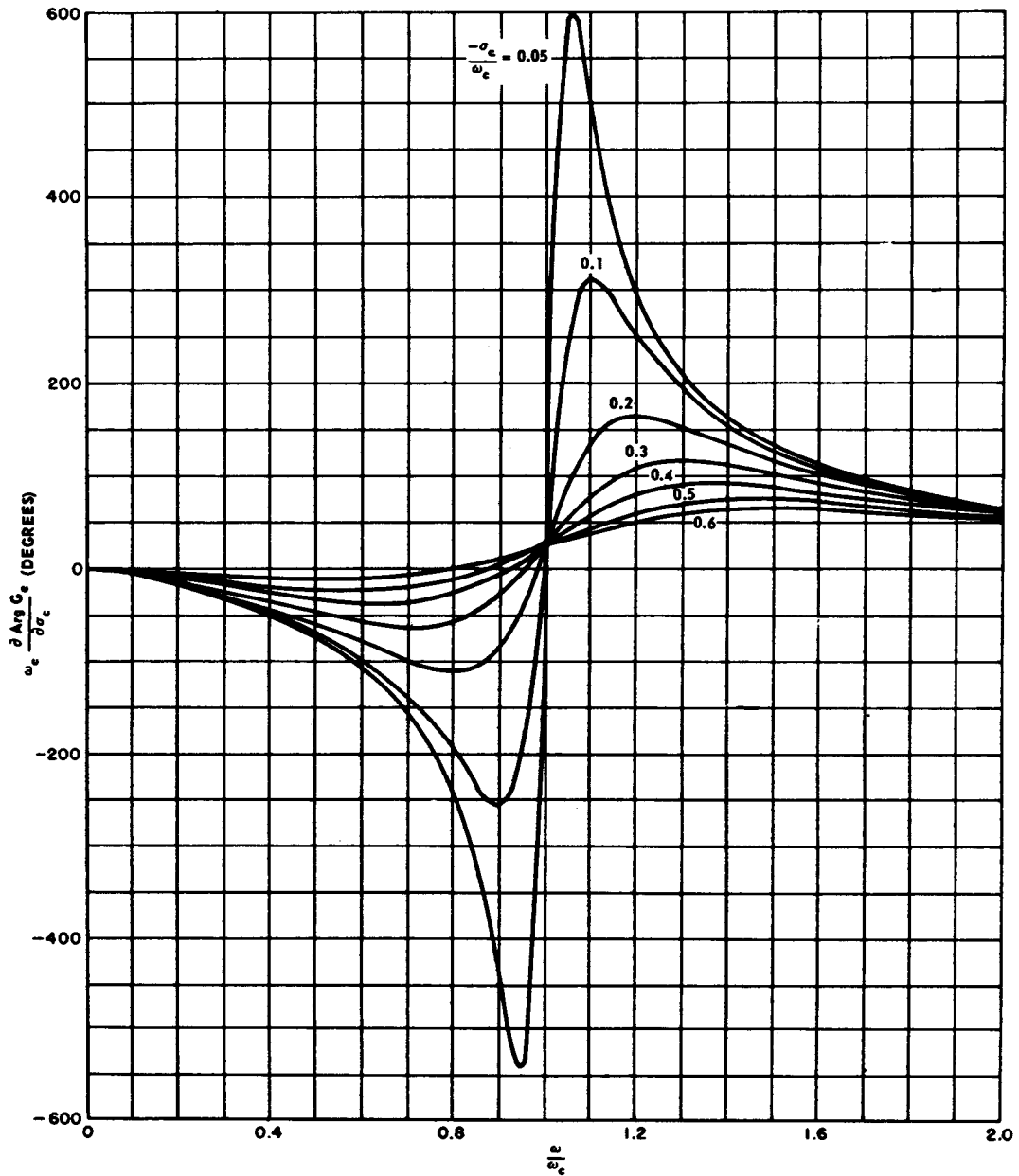


H. DEPENDENCE OF PHASE ON REAL ZERO (OR POLE)

$$G_r = (j\omega - \sigma_r)$$

Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 5 of 10)

THEORY



I. DEPENDENCE OF PHASE ON REAL PART OF ZERO (OR POLE)

$$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$$

Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 6 of 10)

PERFORMANCE EVALUATION

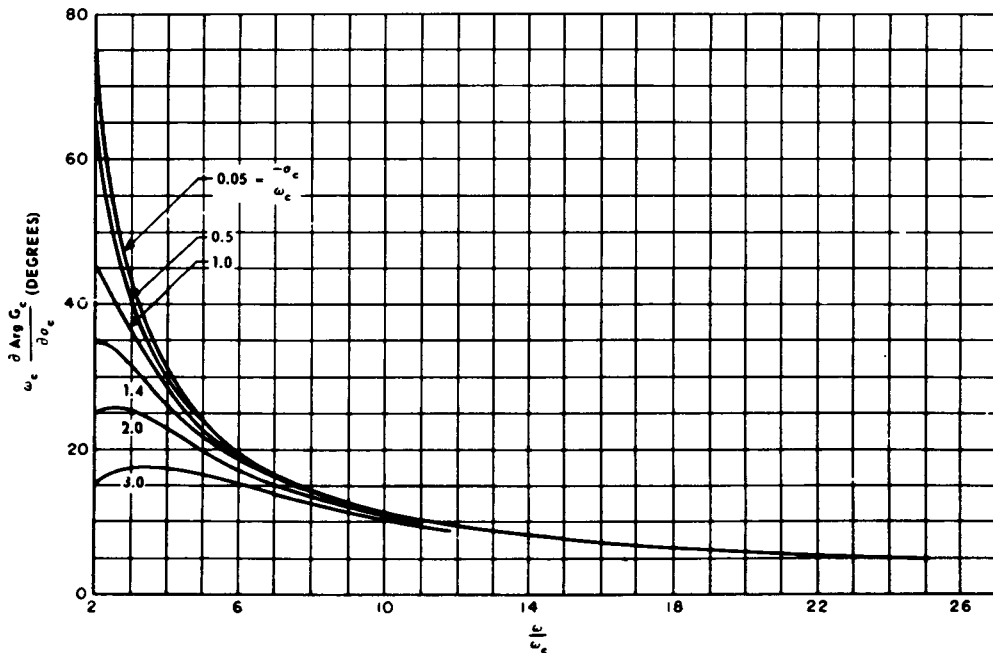
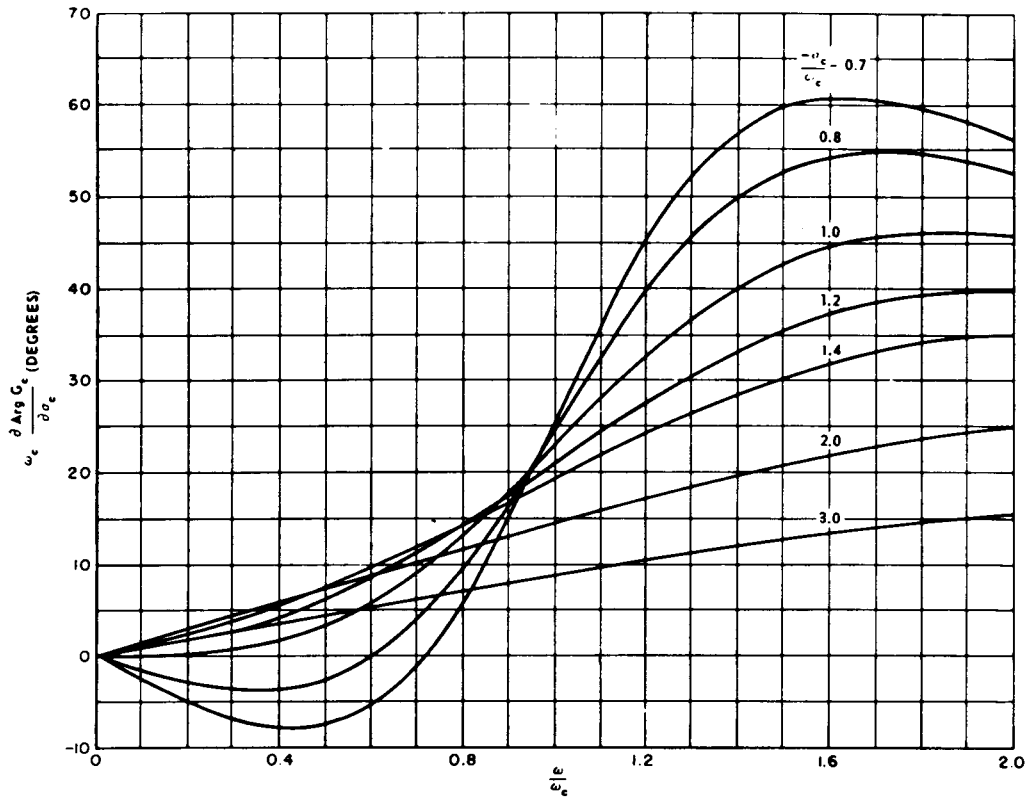
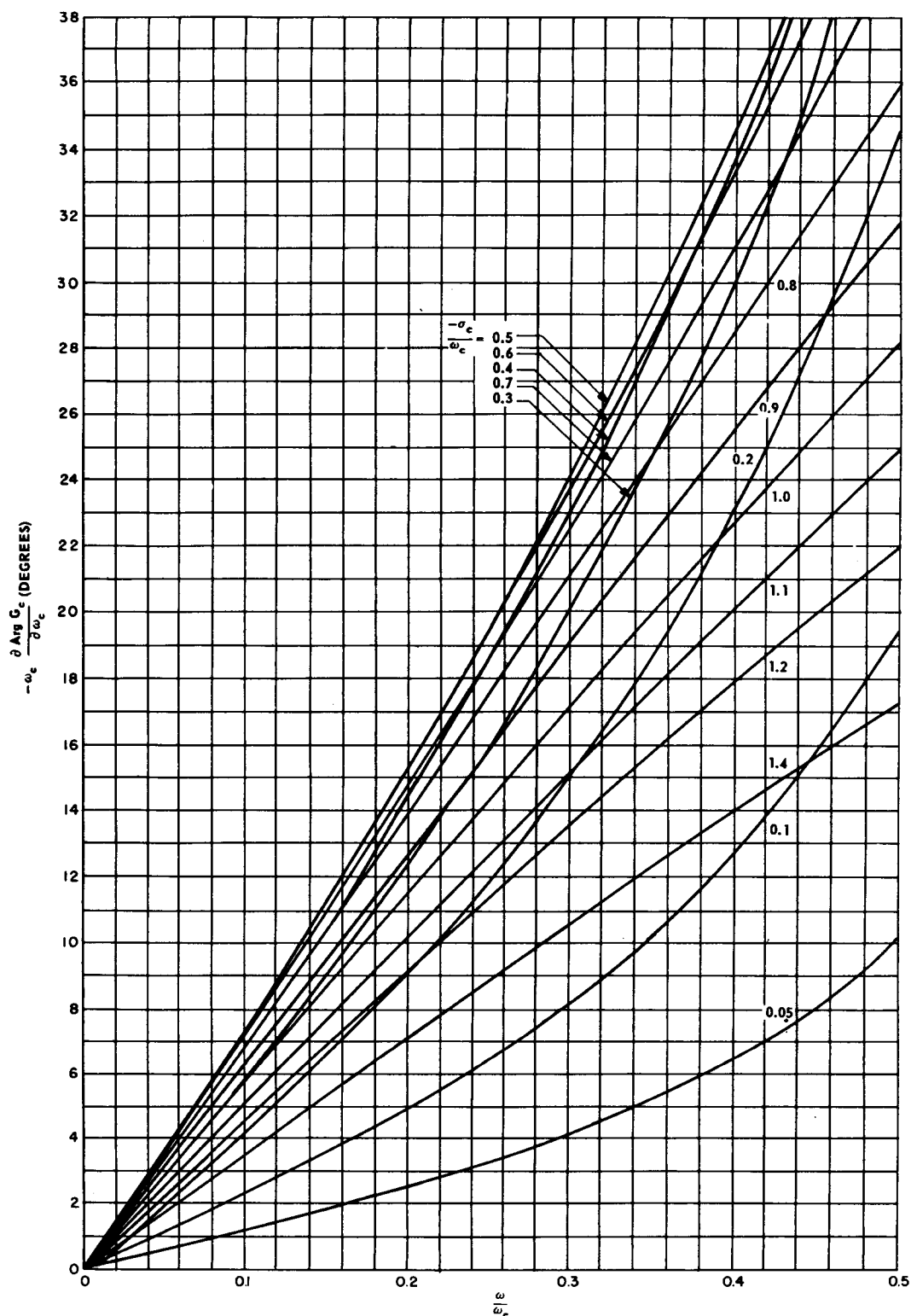


Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 7 of 10)

THEORY

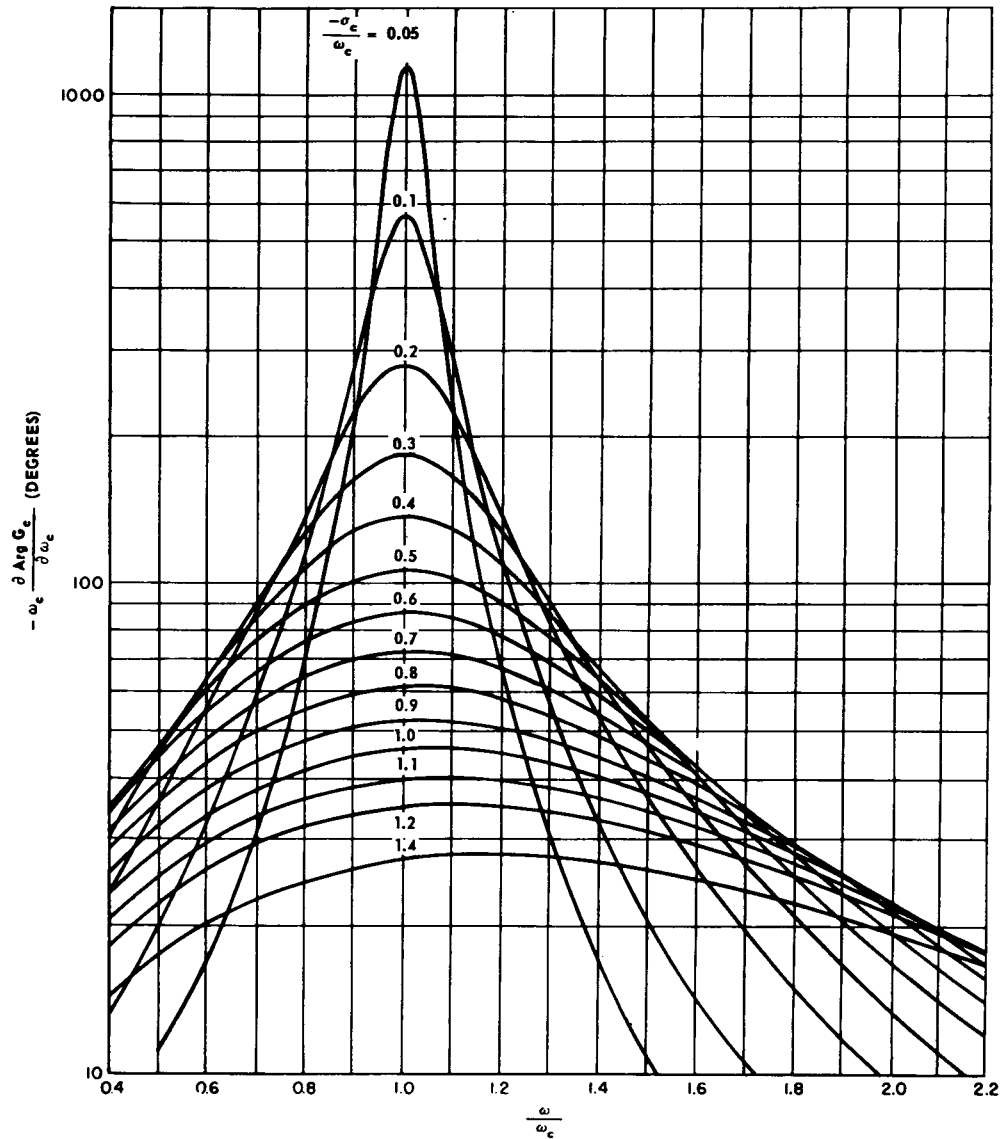


L. DEPENDENCE OF PHASE ON IMAGINARY PART OF ZERO (OR POLE)

$$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$$

Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 8 of 10)

PERFORMANCE EVALUATION



M. DEPENDENCE OF PHASE ON IMAGINARY PART OF ZERO (OR POLE)

$$G_c = [(j\omega)^2 - 2\sigma_c(j\omega) + \sigma_c^2 + \omega_c^2]$$

Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 9 of 10)

THEORY

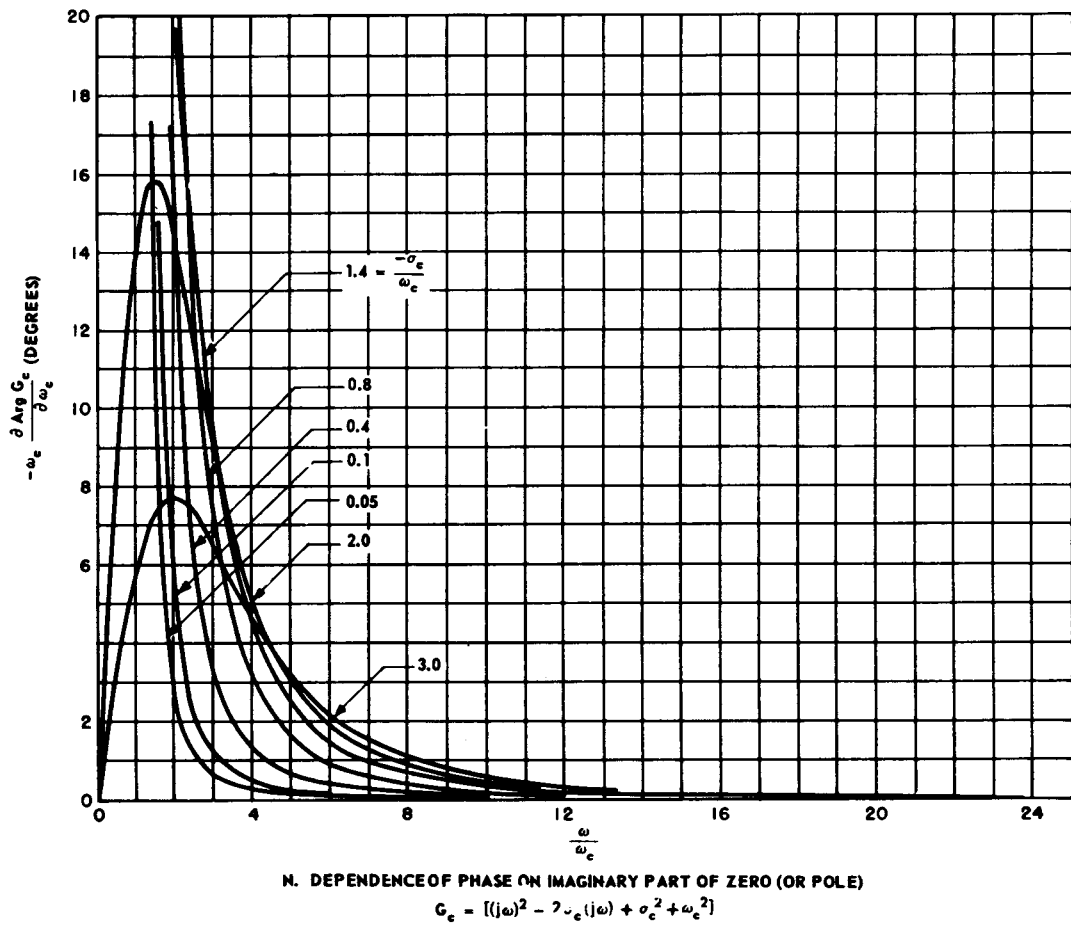


Fig. 7-10 Partial derivatives for Linvill's procedure. (Sheet 10 of 10)

PERFORMANCE EVALUATION

7-1.3 RELATIONS BETWEEN CLOSED-LOOP TRANSIENT RESPONSE AND CLOSED-LOOP POLE-ZERO CONFIGURATION

It is desirable to be able to describe properties of the transient response of a system when one is given the closed-loop pole-zero configuration and vice versa. Usually, the designer is given specifications for some form of the transient response of a system. As a result, the conversion of the transient-response specifications to a desired closed-loop pole-zero configuration is a starting point in many design procedures [see Par. 6-5 and references (3,5,6,19,20,28)]. Since the usual assumption in these design procedures is that the closed-loop performance of the system is primarily controlled by a dominant pair of complex poles (dominant quadratic factor in the denominator), only the characteristics of an underdamped second-order system are presented here.

If the system being examined is a unity-feedback system with a pair of complex-conjugate poles and no closed-loop zeros, the closed-loop transfer function relating output to input is

$$W(s) = \frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (7-12)$$

The error-to-input transfer function is

$$\frac{E(s)}{R(s)} = \frac{s(s + 2\zeta\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (7-13)$$

The open-loop transfer function is

$$G(s) = \frac{C(s)}{E(s)} = \frac{\omega_n^2}{s(s + 2\zeta\omega_n)} \quad (7-14)$$

In these equations,

ω_n = natural frequency

and

ζ = damping ratio.

The magnitude and phase of the closed-loop frequency response $W(j\omega)$ are the second-order quadratic factor curves presented in Par. 5-3. The velocity constant K_v of the system is

$$K_v = \frac{\omega_n}{2\zeta} \quad (7-15)$$

The first three error coefficients are

$$e_0 = 0 \quad (7-16)$$

$$e_1 = \frac{1}{K_v} \quad (7-17)$$

$$e_2 = \frac{1 - 4\zeta^2}{\omega_n^2} \quad (7-18)$$

The error response curves for a *unit-ramp* input are given in Fig. 7-11. Note the steady-state error for a *unit-ramp* input to this system is given by

$$e_{ss} = \frac{2\zeta}{\omega_n} = \frac{1}{K_v} \quad (7-19)$$

The error response curves for a *unit-step* input are given in Fig. 7-12. The output response can be obtained from these curves by subtracting them from unity. *The solution time or settling time t_s* of the step response is the time for the output to reach 98% of its final value or for the error to fall to 2% of its initial value. For the second-order system,

$$t_s = \frac{4}{\zeta\omega_n} \quad (7-20)$$

The output response curves for a *unit-step* input are plotted in Fig. 7-13.

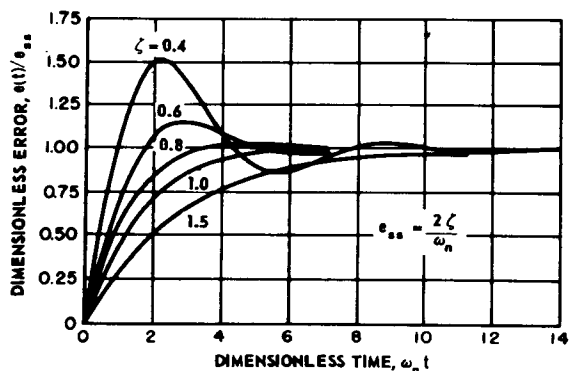


Fig. 7-11 Dimensionless transient error-response curves of a second-order servomechanism to a unit-ramp input.

Adapted with permission from *Principles of Servomechanisms*, by G. S. Brown and D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

THEORY

Quantitative descriptions of the relationships between properties of the transient response and the frequency response of a second-order system will now be given. The resonant frequency of the closed-loop response $W(j\omega)$ is

$$\omega_R = \omega_n \sqrt{1 - 2\zeta^2} \quad (7-21)$$

The magnitude of the resonant peak M_p (see Fig. 7-14) is given by the relation

$$M_p = \frac{1}{2\zeta \sqrt{1 - \zeta^2}} \quad (7-22)$$

The frequency of damped transient oscillation ω_d (damped natural frequency) is

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} \quad (7-23)$$

The time taken to reach the first peak in the output response to a unit-step input is

$$t_{m1} = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}} \quad (7-24)$$

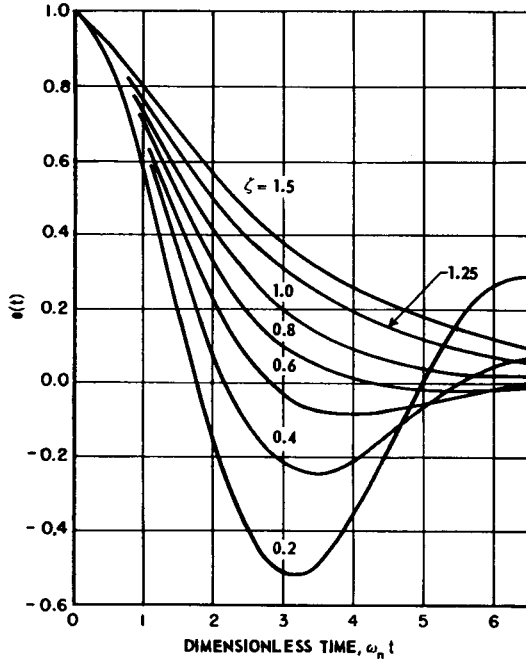


Fig. 7-12 Transient error-response curves of a second-order servomechanism to a unit-step input.

Adapted with permission from *Principles of Servomechanisms*, by G. S. Brown and D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

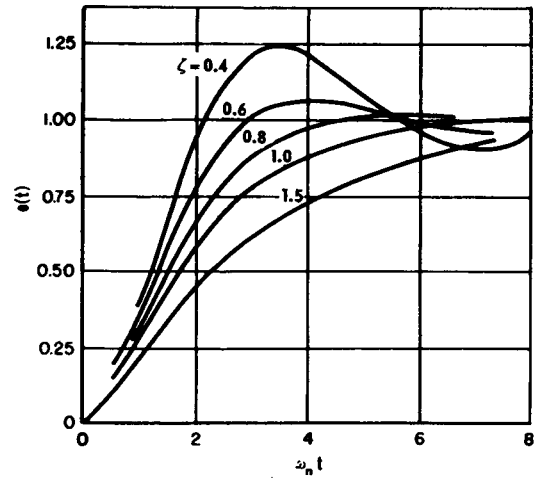


Fig. 7-13 Transient output-response curves of a second-order servomechanism to a unit-step input.

Adapted with permission from *Principles of Servomechanisms*, by G. S. Brown and D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

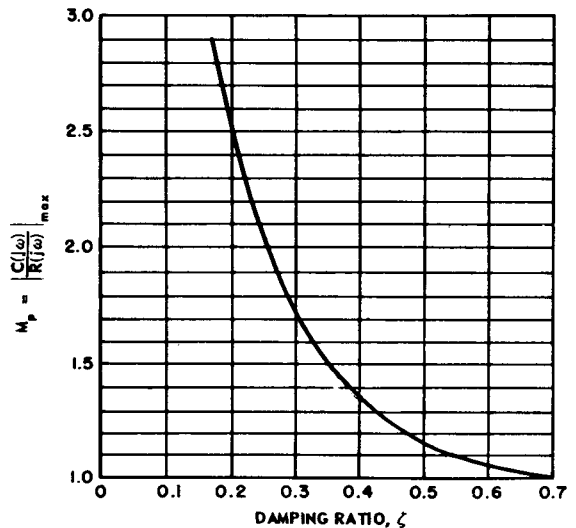


Fig. 7-14 Peak magnitude M_p versus damping ratio ζ for a second-order servomechanism.

Adapted with permission from *Principles of Servomechanisms*, by G. S. Brown and D. P. Campbell, Copyright, 1948, John Wiley & Sons, Inc.

PERFORMANCE EVALUATION

The peak overshoot P_{ov} in the output response to a unit-step input (see Fig. 7-15) is

$$P_{ov} = e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}} \quad (7-25)$$

If bandwidth is defined as the frequency ω_b at which $10 \log_{10} |W(j\omega)|$ is down 1.5 dg from the zero-frequency value, then

$$\omega_b = \omega_n \left[1 - 2\zeta^2 + \sqrt{2 - 4\zeta^2 + 4\zeta^4} \right]^{1/2} \quad (7-26)$$

With the important characteristics of a second-order system described, it is possible to use these characteristics to aid in establishing a desired closed-loop pole-zero configuration from the transient-response specifications.

A few *general* relations between pole-zero configurations and transient-response characteristics are in order. Most closed-loop response functions of unity-feedback systems, $W(s)$, are characterized by a pair of dominant complex poles, one or more dipoles (pole and zero close together), one or more finite zeros, and poles whose magnitudes are much greater (a factor of five or more) than the magnitude

of the dominant pair of poles.^(3,5,6,12,18,19,28) The pertinent relations are as follows:

(a) The addition of a real zero to $W(s)$ tends to increase the overshoot of the output response to a unit-step input, decreasing the rise time and time delay.

(b) The addition of a real pole to $W(s)$ tends to decrease the overshoot of the output response to a unit-step input, increasing the rise time and time delay.

(c) The addition of real poles to $W(s)$ whose magnitudes are much larger than the magnitude of the dominant pole pair has very little effect on the transient response.

(d) The addition of complex poles to $W(s)$ whose magnitudes are much larger than the magnitude of the dominant pole pair has very little effect on the transient response provided the damping ratio of the added poles is not too small.

(e) The addition of a dipole to $W(s)$ has very little effect on the step response of the system but may have a pronounced effect on the steady-state errors of the system.

(f) The excess of poles over zeros for $W(s)$ is equal to or greater than the excess of poles over zeros for the fixed-element transfer function $G_f(s)$.

(g) Most military applications require that $W(0) = 1$. This implies that the open-loop transfer functions $C(s)/E(s) = G(s)$ have at least one pole at the origin.

(h) In any system with one open-loop pole at the origin, the first three error coefficients are

$$e_0 = 0 \quad (7-27)$$

$$e_1 = \frac{1}{K_v} = \sum_{j=1}^n \frac{1}{p_j} - \sum_{j=1}^m \frac{1}{z_j} \quad (7-28)$$

$$e_2 = -\frac{1}{2} \left(\frac{1}{K_v^2} + \sum_{j=1}^n \frac{1}{p_j^2} - \sum_{j=1}^m \frac{1}{z_j^2} \right) \quad (7-29)$$

where $-p_j$ is the j th pole of $W(s)$, $-z_j$ is the j th zero of $W(s)$, and K_v is the velocity constant of the system.

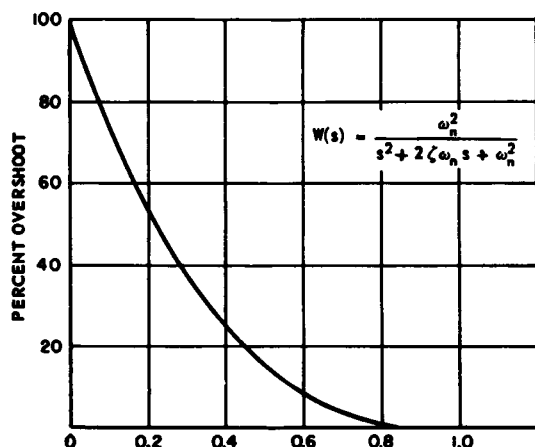


Fig. 7-15 Overshoot variation with ζ .

THEORY

(i) In any system with two open-loop poles at the origin, the first three error coefficients are

$$e_0 = 0 \quad (7-30)$$

$$e_1 = 0 \quad (K_v \rightarrow \infty) \quad (7-31)$$

$$e_2 = -\frac{1}{2} \left(\sum_{j=1}^n \frac{1}{p_j^2} - \sum_{j=1}^m \frac{1}{z_j^2} \right) = -\frac{1}{K_a} \quad (7-32)$$

where K_a is the acceleration constant, $-p_j$ is the j th pole of $W(s)$, and $-z_j$ is the j th zero of $W(s)$.

(j) If the cutoff frequency ω_{co} is defined as the frequency at which the phase of the closed-loop frequency response is -90° and the buildup time t_{bu} is the time for the output response first to cross unity for a step input, then

$$t_{bu} \cong \frac{\pi}{\omega_{co}} \quad (7-33)$$

(k) If the rise time t_r is defined as the time required for the output response to a unit-step input to go from 10% to 90% of its final value and ω_b is the bandwidth as defined immediately above Eq. (7-26), then for a response with less than 10% overshoot,

$$\frac{\omega_b t_r}{2\pi} \cong 0.30 \text{ to } 0.45 \quad (7-34)$$

(l) If the delay time t_d is defined as the time for the output response to a unit step to reach 50% of its final value, then

$$t_d \cong \frac{1}{K_v} \quad (7-35)$$

(m) If the rise time t_r is as defined in (k),

$$\frac{1}{2\pi} t_r^2 \cong -\left(2e_2 + \frac{1}{K_v^2} \right) \quad (7-36)$$

where e_2 is the second error coefficient and K_v is the velocity constant.

(n) If the settling time (solution time) t_s is the time for the output response to a unit-step input to reach 98% of its final value, then

$$t_s \cong 3t_{bu} \text{ to } 5t_{bu} \quad (7-37)$$

From the characteristics of the second-order system and the general relations (a) through (n) of the preceding paragraph, the conversion of time-domain specifications to a closed-loop pole-zero configuration becomes a fairly straightforward matter. Truxal⁽²⁸⁾ presents a very good description of a typical procedure.

Example. The specifications for a servo-mechanism are as follows:

$$(a) \quad G_f(s) = \frac{K}{s(s+a)}$$

(b) The bandwidth ω_b of the closed-loop response shall be less than 50 rad/sec, and the output response of the system to a unit-step input shall have an overshoot less than 5% of the final value.

$$(c) \quad K_v \geq 50 \text{ sec}^{-1}$$

$$(d) \quad |e_2| \leq 0.01 \text{ sec}^2$$

Find a closed-loop pole-zero configuration that satisfies these specifications.

Solution. If the system is initially approximated by a second-order response with no zeros, Fig. 7-15 shows that $\zeta \geq 0.7$ for a peak overshoot $P_{ov} \leq 5\%$. For $\zeta = 0.7$, Eq. (7-26) yields $\omega_b = \omega_n$. Therefore, $\omega_n \leq 50$ rad/sec. The dominant pole pair is thus placed at $s = -35 \pm 35j$ corresponding to $\omega_n = 49.5$ rad/sec and $\zeta = 0.707$. Now using Eq. (7-15), we find that $K_v = 35 \text{ sec}^{-1}$, which is too small a value. To increase K_v , a dipole will be added. The pole of the dipole must not be set at too low a frequency or else an excessively long tail in the transient will occur. The magnitude of the pole of the dipole will therefore be chosen to be one-tenth the real part of the dominant poles. This corresponds to a pole at -3.5 . To determine the location of the zero of the dipole, Eq. (7-28) is used. At this point, the approximate closed-loop response is

$$W(s) \cong \frac{\frac{p_1}{z_1} \omega_n^2 (s + z_1)}{(s^2 + 2\zeta\omega_n s + \omega_n^2)(s + p_1)}$$

where $p_1 = +3.5$, $\zeta = 0.7$, $\omega_n = 49.5$, and z_1 is to be determined. Using Eq. (7-23),

$$\frac{1}{K_v} = \frac{1}{p_1} + \frac{1}{\zeta\omega_n + j\omega_n\sqrt{1-\zeta^2}} + \frac{1}{\zeta\omega_n - j\omega_n\sqrt{1-\zeta^2}} - \frac{1}{z_1}$$

or,

$$\frac{1}{z_1} = \frac{1}{p_1} + \frac{2\zeta}{\omega_n} - \frac{1}{K_v}$$

Therefore, $z_1 = 3.40$. The desired pole-zero configuration for $W(s)$ is given by

$$W(s) \cong \frac{2520(s + 3.4)}{(s^2 + 70s + 2450)(s + 3.5)}$$

The second error coefficient of this system can be found from Eq. (7-29). Thus,

$$e_2 = -\frac{1}{2} \left[\frac{1}{K_v^2} + \frac{1}{p_1^2} + \left(\frac{1}{\zeta\omega_n + j\sqrt{1-\zeta^2}} \right)^2 + \left(\frac{1}{\zeta\omega_n - j\sqrt{1-\zeta^2}} \right)^2 - \frac{1}{z_1^2} \right]$$

Evaluating this expression, it is found that $e_2 = 2 \times 10^{-8} \text{ sec}^2$, which is well within specifications. All the specifications should be checked at this point to insure that the system behaves as desired. The example above has been carried out far enough to demonstrate the basic ideas involved in finding a closed-loop pole-zero configuration that satisfies the given specifications.

7-1.4 RELATIONS BETWEEN OPEN-LOOP FREQUENCY RESPONSE AND CLOSED-TRANSIENT RESPONSE (4,7,8,11,12,25,26)

Since most of the design techniques discussed in Pars. 6-2, 6-3, and 6-4 involve considerations of the open-loop frequency response $C(j\omega)/E(j\omega) = G(j\omega)$, methods for relating the open-loop frequency response to the closed-loop transient response will be presented here.

Harris et al.⁽⁷⁾ present an approximate technique for determining the error response $e(t)$ to a transformable input $r(t)$. If ω_c is defined as the frequency at which the open-loop asymptotes cross 0 dg (asymptote crossover frequency; see Fig. 7-16), this method

assumes that ω_c occurs in a region where the slope of the asymptote is -10 dg/dec . In general, the shape of the open-loop asymptote for frequencies greater than ω_c has little effect on the transient response of the system.

The reciprocal error-to-input transfer function $R(s)/E(s)$ can be found from the open-loop response $C(s)/E(s)$ by using the relation

$$\frac{R(s)}{E(s)} = \frac{C(s)}{E(s)} + 1 \quad (7-38)$$

Since the open-loop asymptote function $C(s)/E(s)$ is almost always a monotonically decreasing function of frequency, the asymptote crossover frequency ω_c divides the frequency scale into two regions:

$$\frac{R(s)}{E(s)} \cong \frac{C(s)}{E(s)} \text{ for } \omega \ll \omega_c \quad (7-39)$$

$$\frac{R(s)}{E(s)} \cong 1 \text{ for } \omega \gg \omega_c \quad (7-40)$$

The procedure for finding $e(t)$ is as follows:

(a) From $[R(s)/E(s)]_{\text{approximate}}$ by using all factors of $C(s)/E(s)$ corresponding to poles and zeros of $C(s)/E(s)$ with magnitudes (break frequencies) less than ω_c . Delete all other factors of $C(s)/E(s)$ and add a numerator factor equal to $\left(1 + \frac{s}{\omega_c}\right)$.

(b) $[E(s)/R(s)]_{\text{approximate}}$ is the reciprocal of $[R(s)/E(s)]_{\text{approximate}}$. From the transform $R(s)$ of the input $r(t)$ and the approximate error-to-input transfer function, find the first

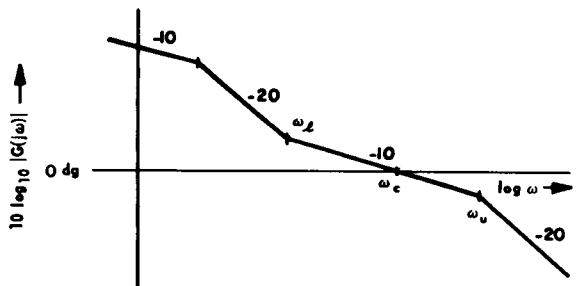


Fig. 7-16 Typical open-loop asymptote function.

THEORY

approximation to $e(t)$ by performing the inverse Laplace transformation of the function $[E(s)/R(s)]_{\text{approximate}} \times R(s)$ and plotting this time function.

(c) Find the correction ratio,

$$\rho = \left(\frac{[R(s)/E(s)]_{\text{approximate}}}{[R(s)/E(s)]_{\text{exact}}} \right) \quad (7-41)$$

$s = j\omega_c$

(d) The ends ω_l and ω_u of the -10 dg/dec slope region which fixes ω_c are called the lower-and upper-corner (or break) frequencies of the -10 dg/dec region (see Fig. 7-16). The plot of the first approximation to $e(t)$ found in (b) is multiplied by the correction ratio ρ in the time interval

$$\frac{1}{\omega_u} < t < \frac{1}{\omega_l} \quad (7-42)$$

and the resulting curves are joined smoothly in the regions $t = 1/\omega_u$ and $t = 1/\omega_l$. This method works best if the -10 dg/dec slope region is fairly long ($\omega_u/\omega_l \cong 8$) and if the closed-loop M_p is close to unity.

Chestnut and Mayer ^(8,26) present a series of charts that can be used to determine the properties of a unity-feedback system from the asymptotes of the open-loop frequency response. These charts utilize the following terminology (see Fig. 7-17) :

$\left. \frac{C}{R} \right|_m \triangleq M_p$, the maximum ratio of closed-loop frequency response

$\left. \frac{C}{R} \right|_p \triangleq$ the peak value of the ratio of controlled variable (output) to reference variable (input) for a step input

$\frac{\omega_m}{\omega_c} \triangleq$ the ratio of the frequency ω_m at which $\left. \frac{C}{R} \right|_m$ occurs to the frequency ω_c at which the straight-line approximation (asymptote) of the open-loop response is 0 decibels. (Note: 2 decibels = 1 decibel.)

$\frac{\omega_l}{\omega_c} \triangleq$ the ratio of ω_l , the lowest frequency of oscillation for a step input, to ω_c , the frequency at which the open-loop asymptote crosses 0 db (decibels).

$\omega_c t_p \triangleq$ the asymptote crossover frequency ω_c times t_p , the response time from the start of the step function until $\left. \frac{C}{R} \right|_p$ occurs.

$\omega_c t_s \triangleq$ the asymptote crossover frequency ω_c times t_s , the settling time from the start of the step function until the output continues to differ from the input by less than 5%.

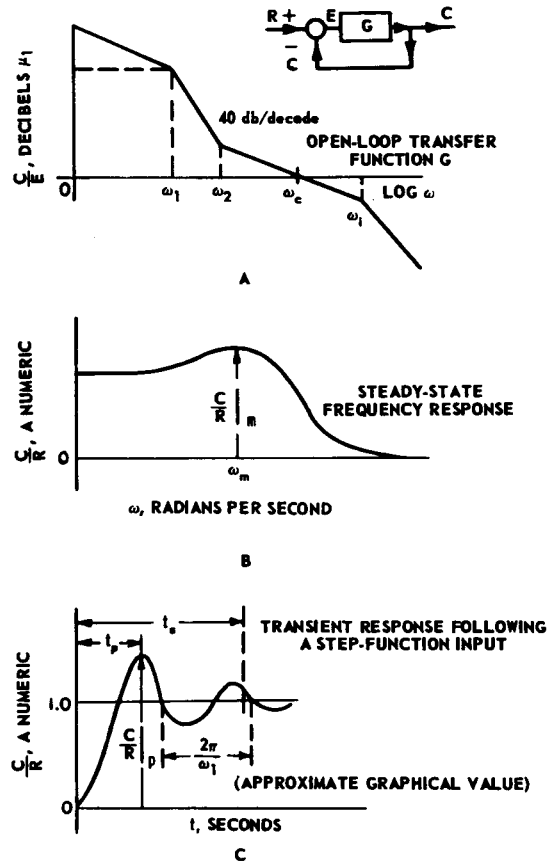


Fig. 7-17 Sketches showing nomenclature used to describe various characteristics of servomechanism performance.

Reprinted with permission from *Servomechanisms and Regulating System Design*, Volume I, by H. Chestnut and R. W. Mayer, Copyright, 1951, John Wiley & Sons, Inc.

PERFORMANCE EVALUATION

Additional definitions are given in Fig. 7-17. The charts correlating the quantities defined here are presented in Fig. 7-18 (sheets 1-18). It should be noted that these charts can be used either for analysis or for synthesis.

Biernson⁽¹¹⁾ presents an excellent method for determining the closed-loop poles of a system from the open-loop frequency response. If it is assumed that the asymptote crossover frequency ω_c of the open-loop frequency response occurs in (or near) a frequency region where the slope of the asymptote is -10dg/dec , then the following relations hold if $|G(j\omega)|$ is a monotonically decreasing function:

$$|G(j\omega)| \cong 1 \text{ for } \omega \cong \omega_c \quad (7-43)$$

$$|G(j\omega)| \gg 1 \text{ for } \omega < \omega_c \text{ (low-frequency range)} \quad (7-44)$$

$$|G(j\omega)| \ll 1 \text{ for } \omega > \omega_c \text{ (high-frequency range)} \quad (7-45)$$

The first approximation to the location of the poles of the closed-loop transfer function $C(s)/R(s)$ is obtained from the following:

(a) The zeros of $G(s)$ whose magnitudes are less than ω_c (low-frequency zeros)

(b) The poles of $G(s)$ whose magnitudes are greater than ω_c (high-frequency poles)

(c) A pole at $s = -\omega_c$

For real or complex closed-loop poles which are far from ω_c in magnitude, the shift from the first approximation of these poles to their actual location can be calculated by a successive-approximation method which converges more rapidly the further the poles are from ω_c . If a closed-loop pole is approximated by a low-frequency zero of $G(s)$, then the true location of the closed-loop pole s_1 can be determined by successively evaluating

$$(\delta_a)^n \cong \left[\frac{(s - s_a)^n}{G(s)} \right]_{s=s_1} \quad \text{for } |s_a| < \omega_c \text{ and } s_1 \cong s_a \quad (7-46)$$

where n is the order of the open-loop zero, s_a is the location of the open-loop zero, and δ_a is the shift from the open-loop zero to the closed-loop pole, i.e., $\delta_a = s_1 - s_a$.

If a closed-loop pole is approximated by a high-frequency pole of $G(s)$, then the true location of the closed-loop pole s_2 can be determined by evaluating

$$(\delta_b)^n \cong -[(s - s_b)^n G(s)]_{s=s_2} \quad \text{for } |s_b| > \omega_c \text{ and } s_2 \cong s_b \quad (7-47)$$

where n is the order of the open-loop pole, s_a is the location of the open-loop pole, and δ_b is the shift from the open-loop pole to the closed-loop pole, i.e., $\delta_b = s_2 - s_b$.

For closed-loop poles near ω_c , a graphical procedure⁽¹¹⁾ is recommended since the convergence of the numerical method employing Eqs. (7-46) and (7-47) is either slow or nonexistent. The graphical procedure involves plots of $G(s)$ for $s = -\sigma + j\omega$ (along axes other than the imaginary axis). Because the graphical procedure tends to be somewhat lengthy, it will not be given here.

Example. The open-loop transfer function of a unity-feedback system is given by

$$G(s) = \frac{K_o(s + \omega_2)(s + \omega_3)}{s(s + \omega_1)^2(s + \omega_4)^2} \quad (7-48)$$

where

$$\omega_1 = 0.04$$

$$\omega_2 = 0.2$$

$$\omega_3 = 1$$

$$\omega_4 = 16$$

K_o is a proportionality constant whose value is to be determined. The asymptotes of this function are sketched in Fig. 7-19. The crossover frequency ω_c is chosen as the geometric mean of ω_3 and ω_4 since this particular choice tends to produce the lowest closed-loop M_p . Near ω_c , the asymptote is given by

$$\text{Asymp } |G| = \left| \frac{K_o}{s\omega_4^2} \right|$$

At the crossover frequency, ω_c , therefore,

$$\frac{K_o}{\omega_c \omega_4^2} = 1$$

or

$$K_o = \omega_c \omega_4^2 = \sqrt{\omega_3 \omega_4} \omega_4^2 = 4\omega_4^2$$

THEORY

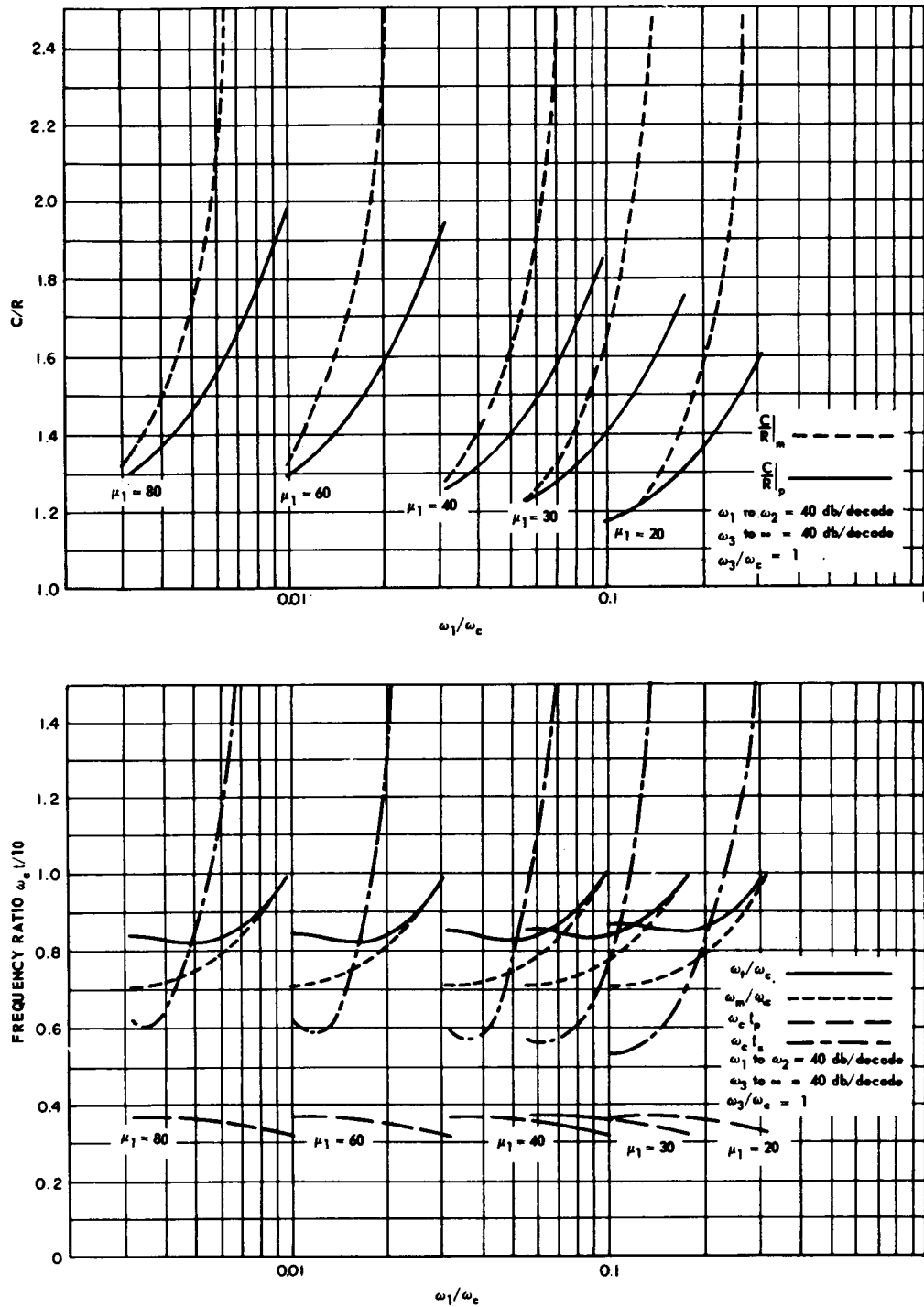


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 1 of 18)

Reprinted with permission from *Servomechanisms and Regulating System Design*, Volume I, by H. Chestnut and R. W. Mayer, Copyright, 1951, John Wiley & Sons, Inc.

PERFORMANCE EVALUATION

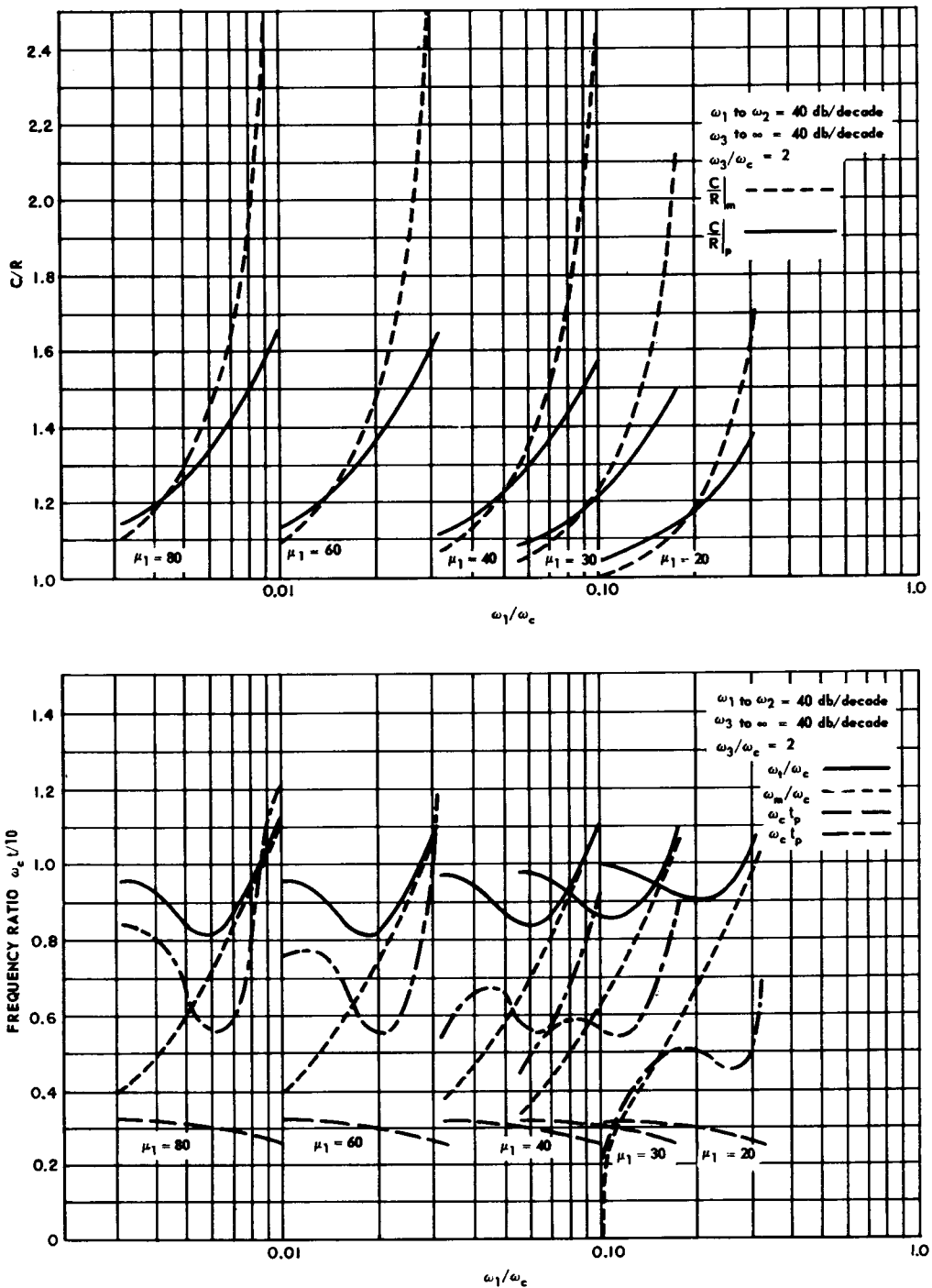


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 2 of 18)

THEORY

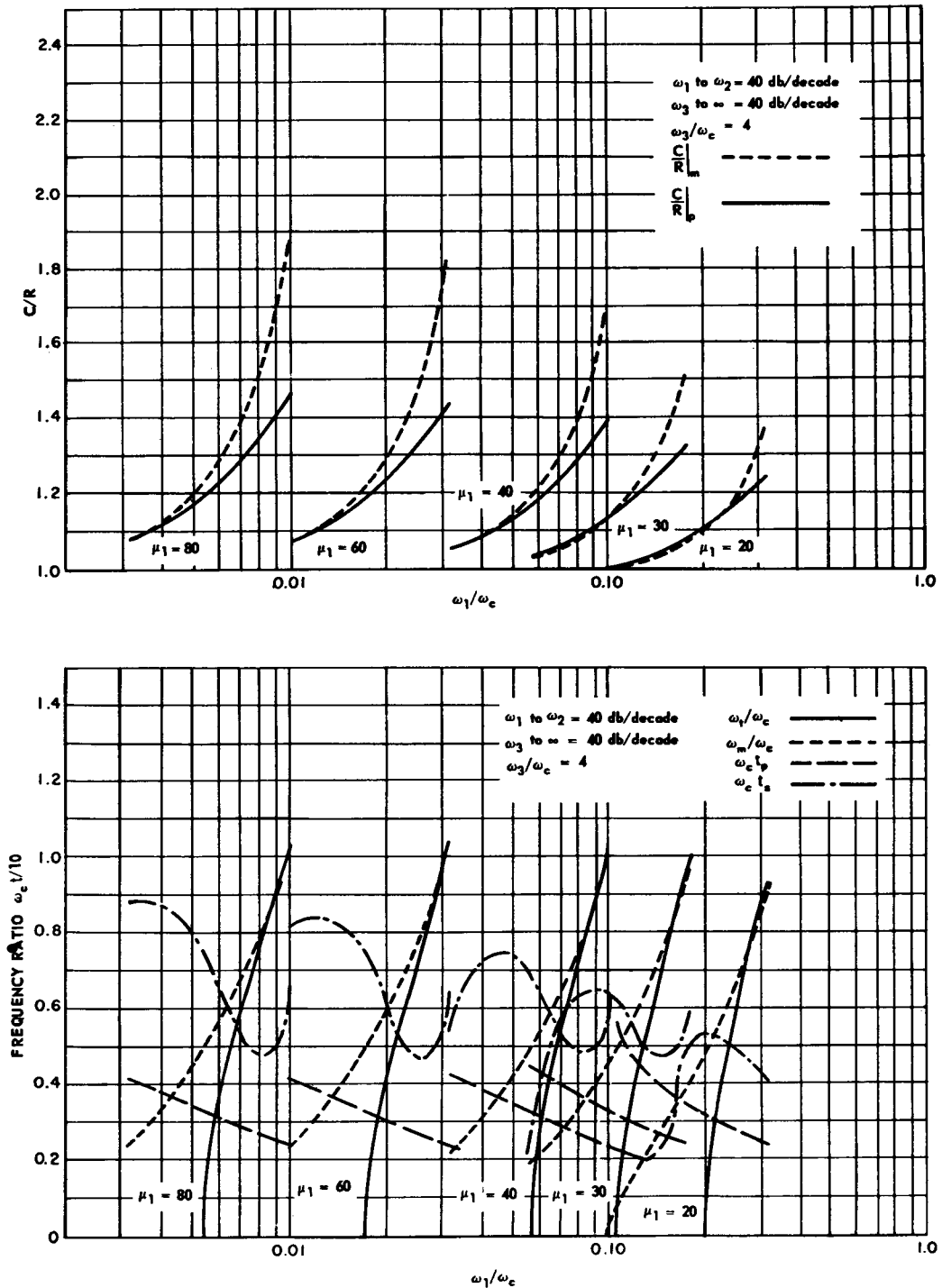


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 3 of 18)

PERFORMANCE EVALUATION

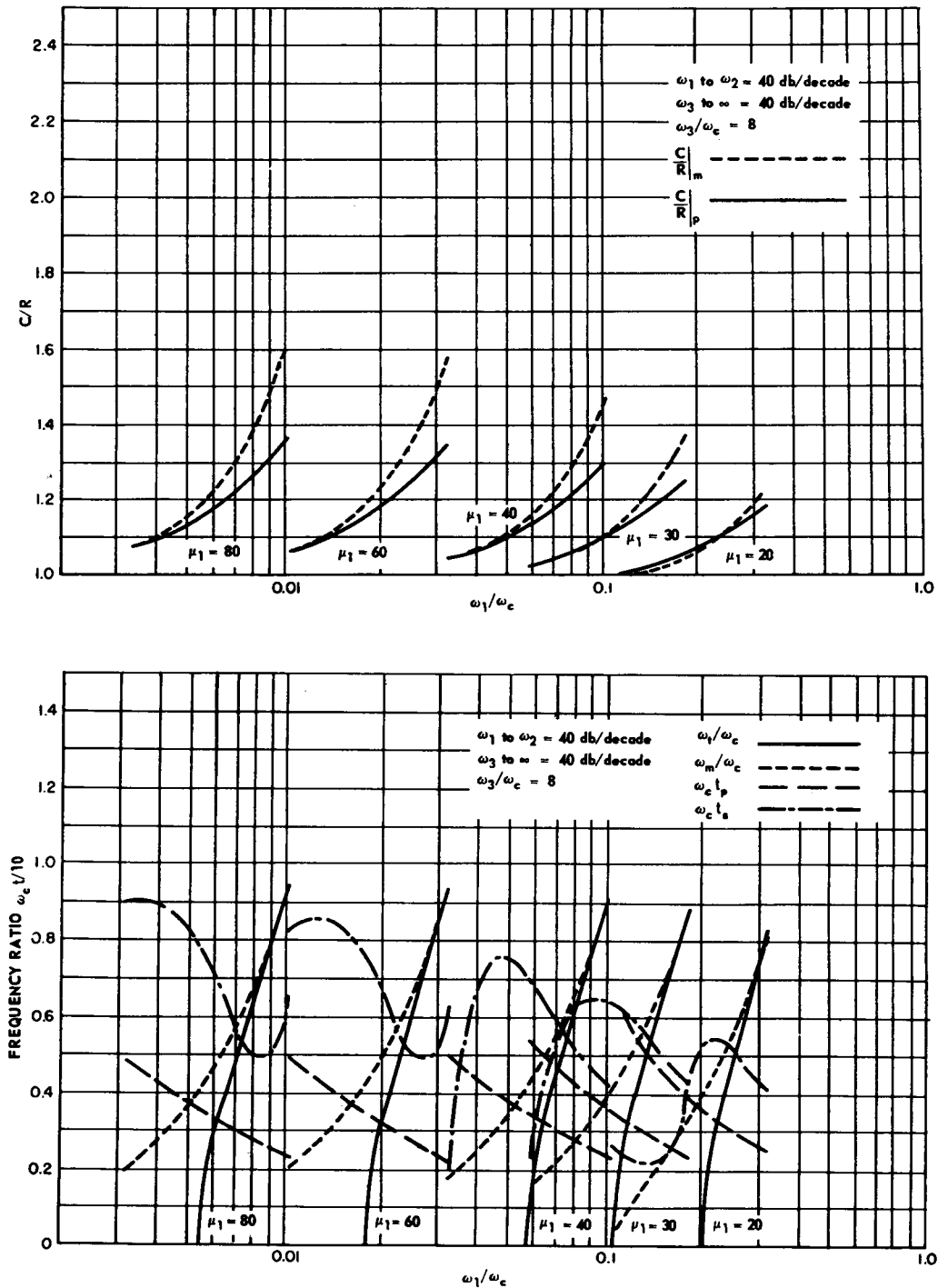


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 4 of 18)

THEORY

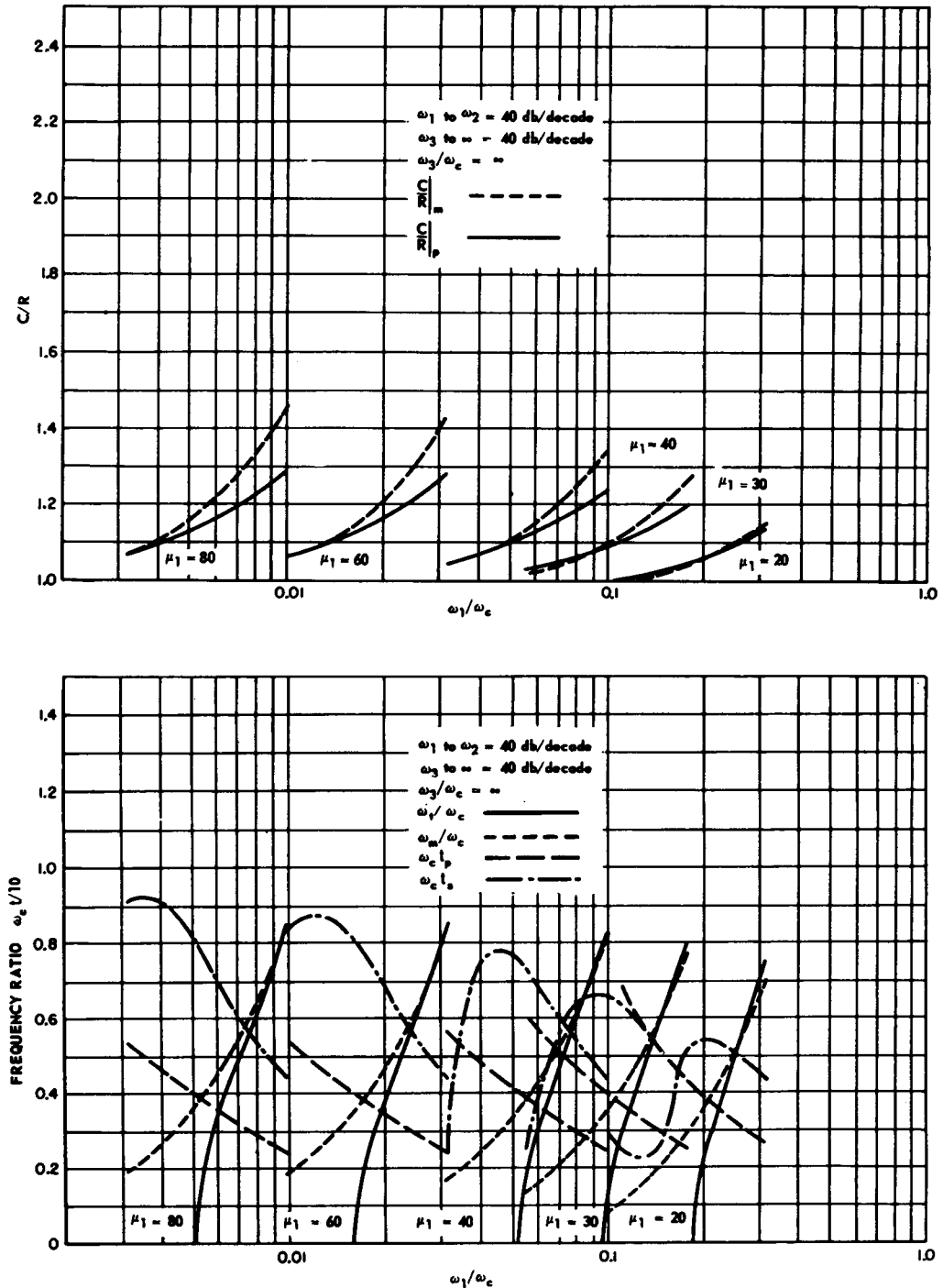


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 5 of 18)

PERFORMANCE EVALUATION

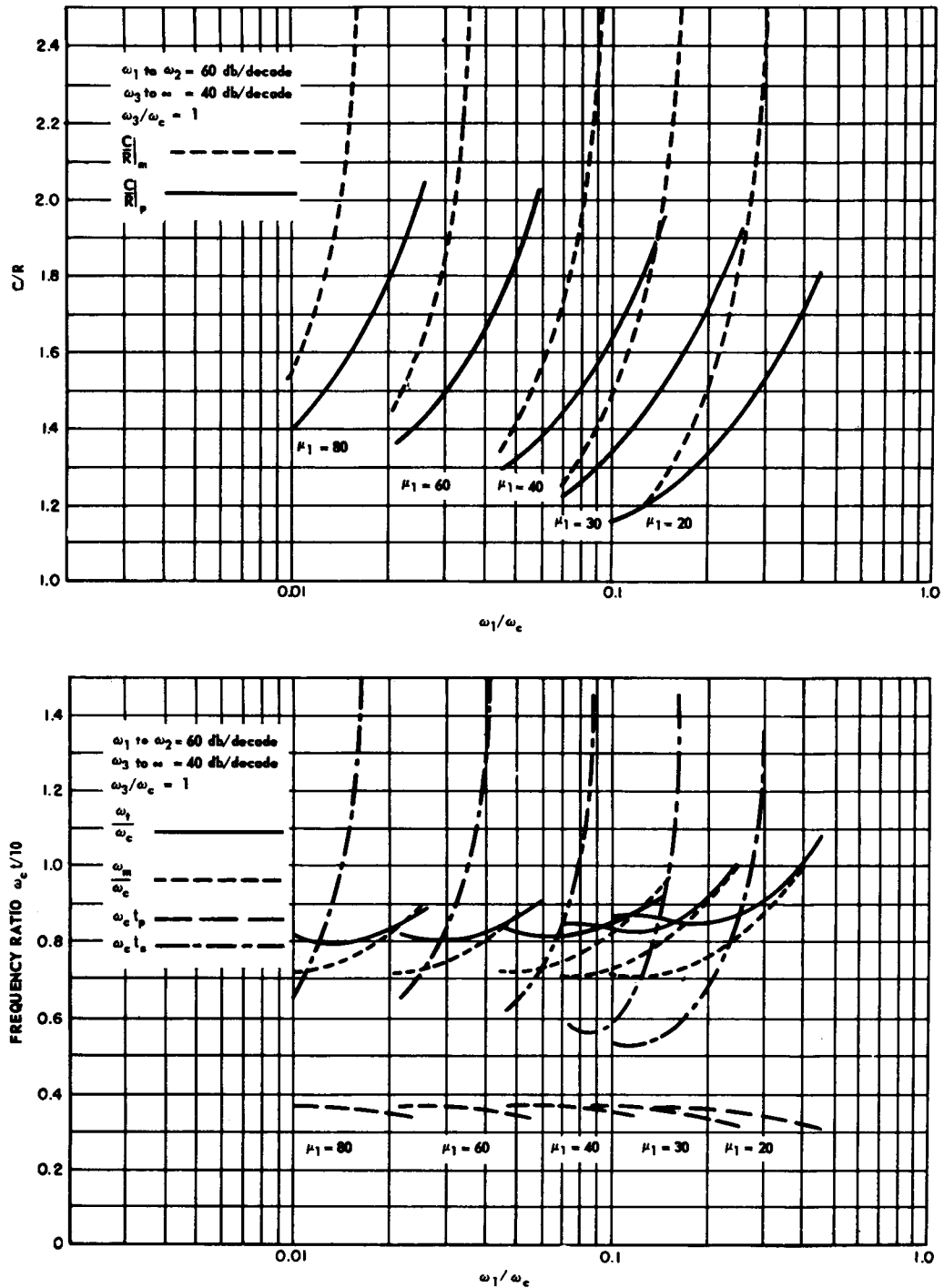


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 6 of 18)

THEORY

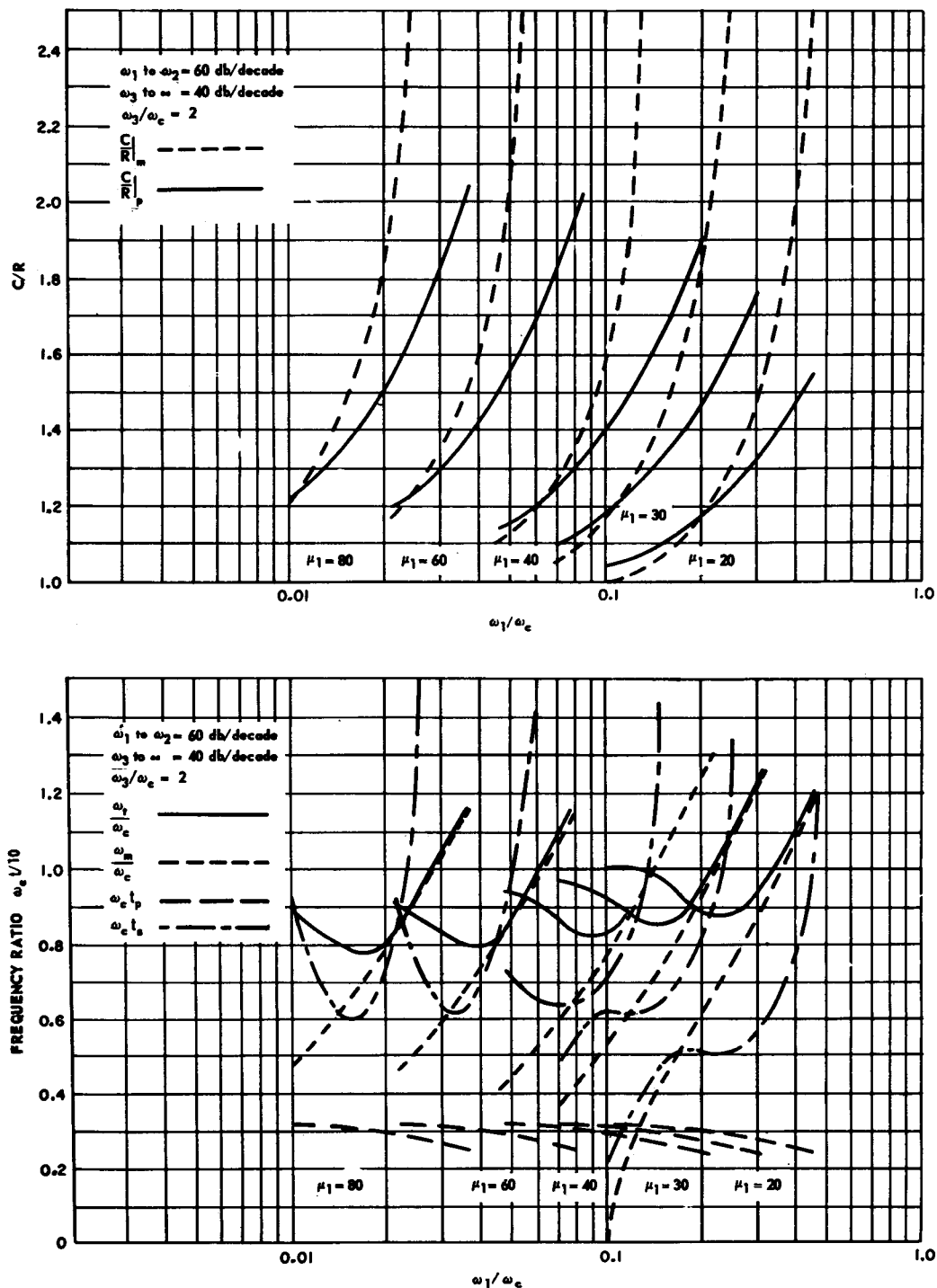


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 7 of 18)

PERFORMANCE EVALUATION

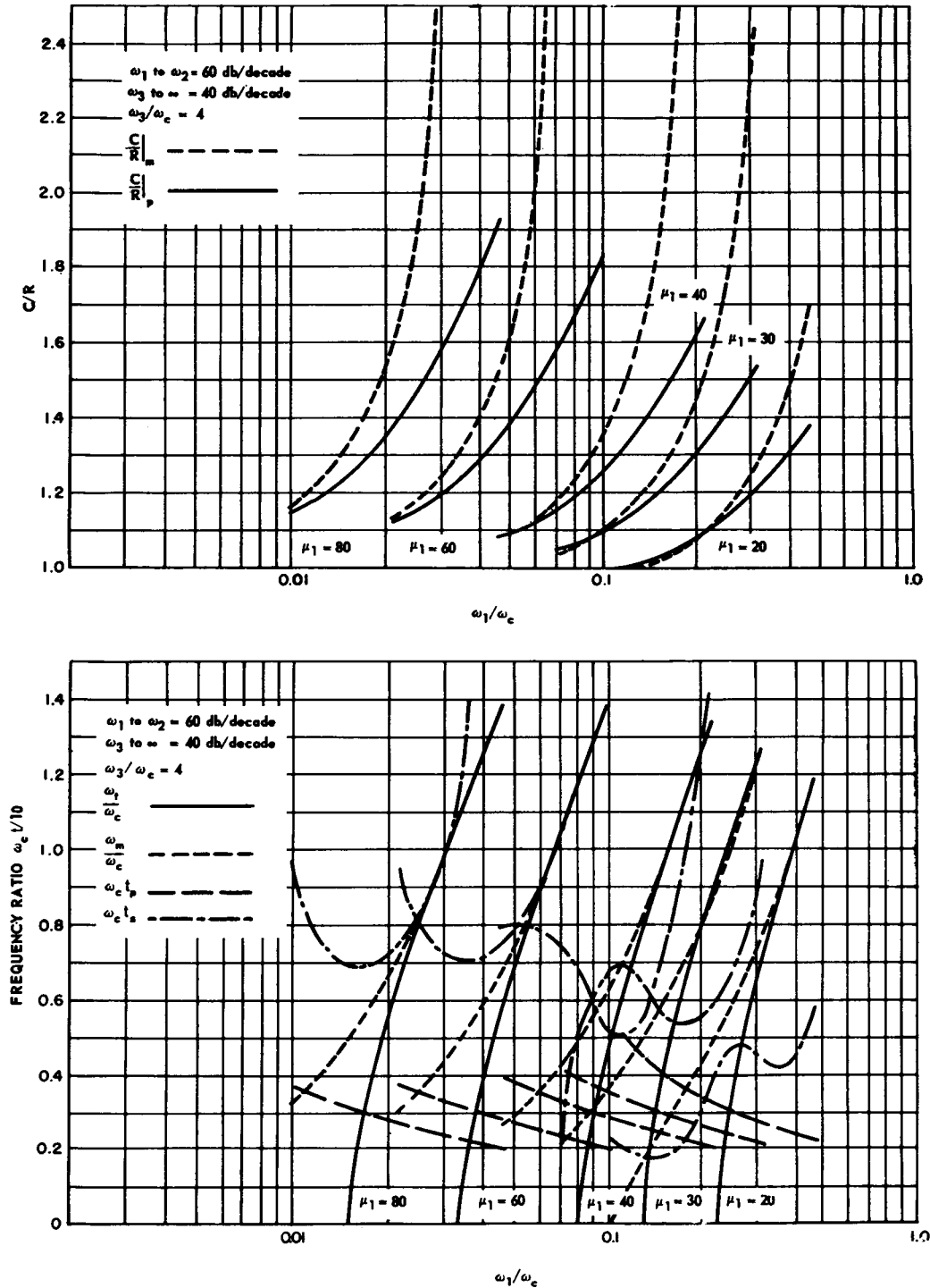


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 8 of 18)

THEORY

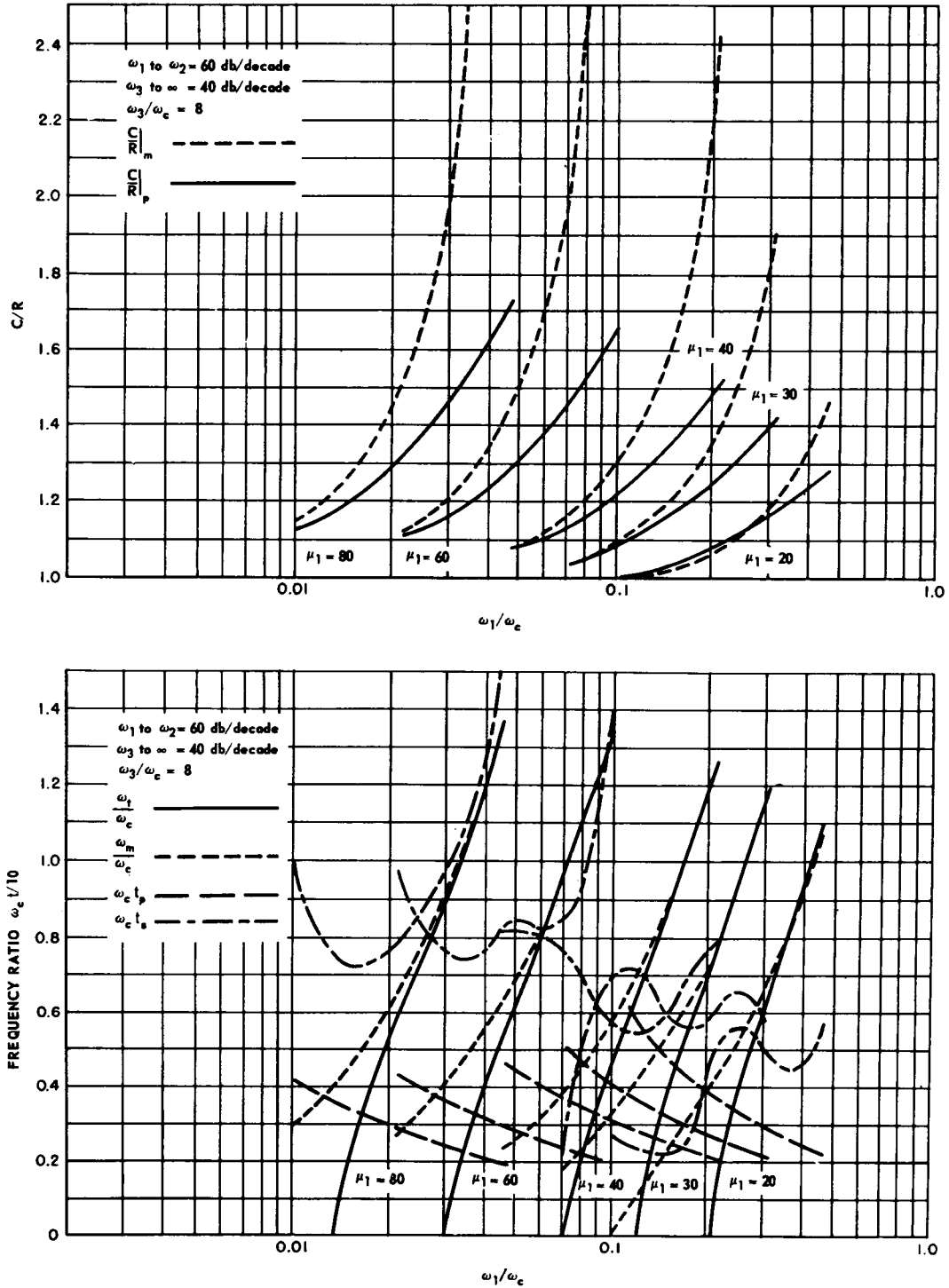


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 9 of 18)

PERFORMANCE EVALUATION

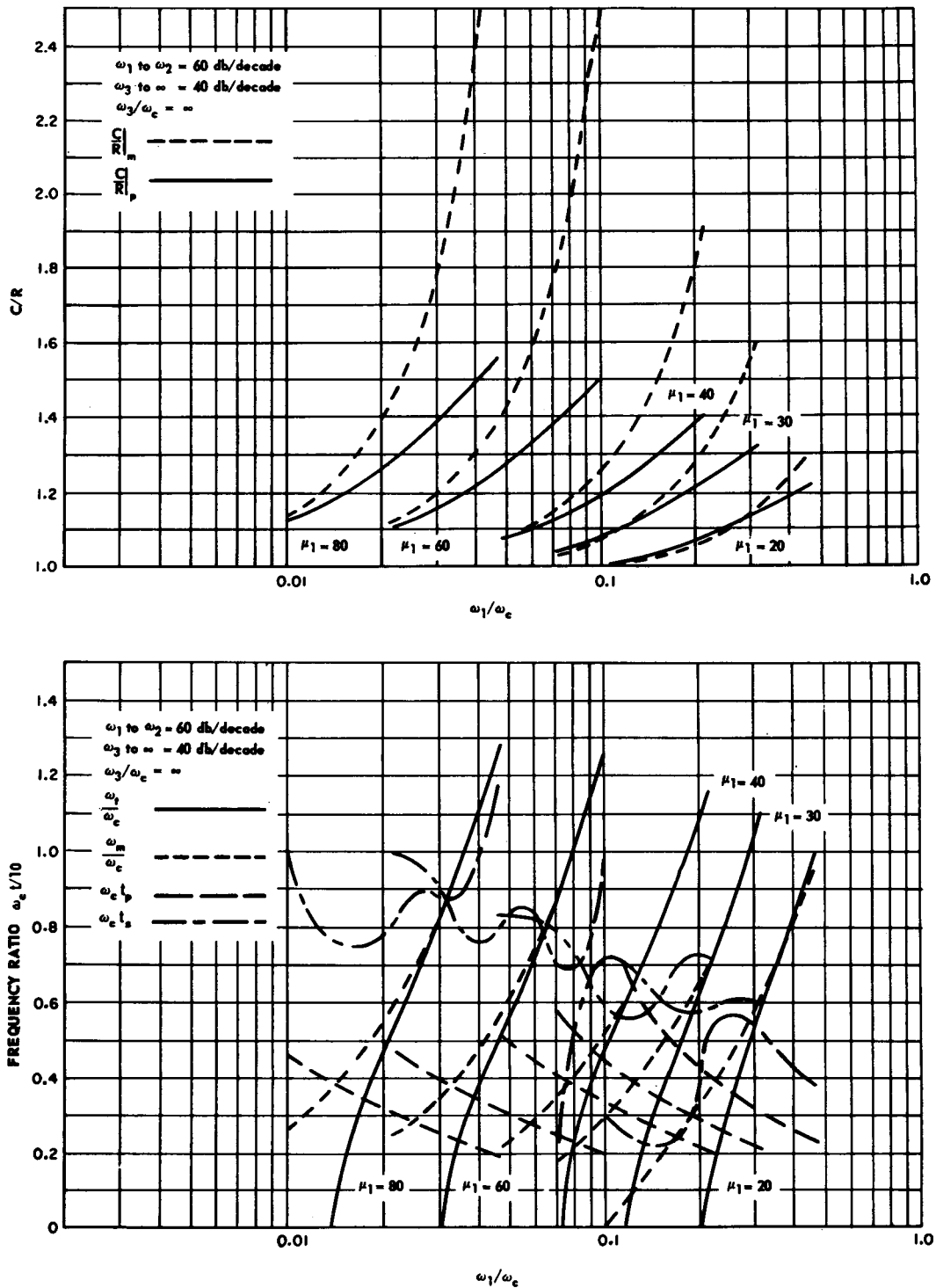


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 10 of 18)

THEORY

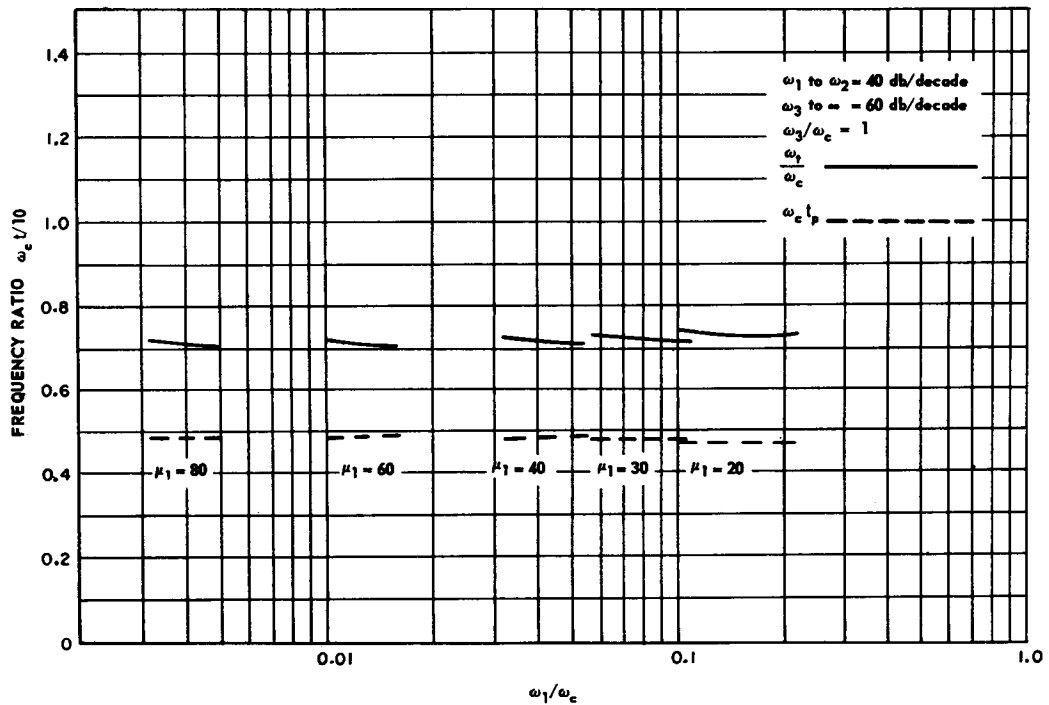
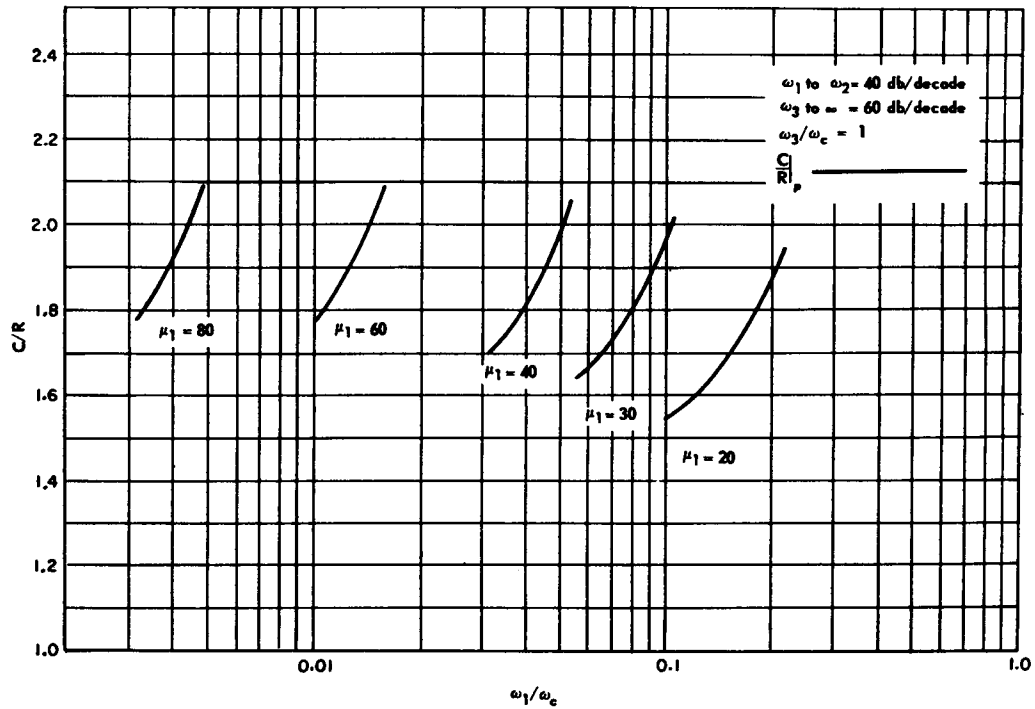


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 11 of 18)

PERFORMANCE EVALUATION

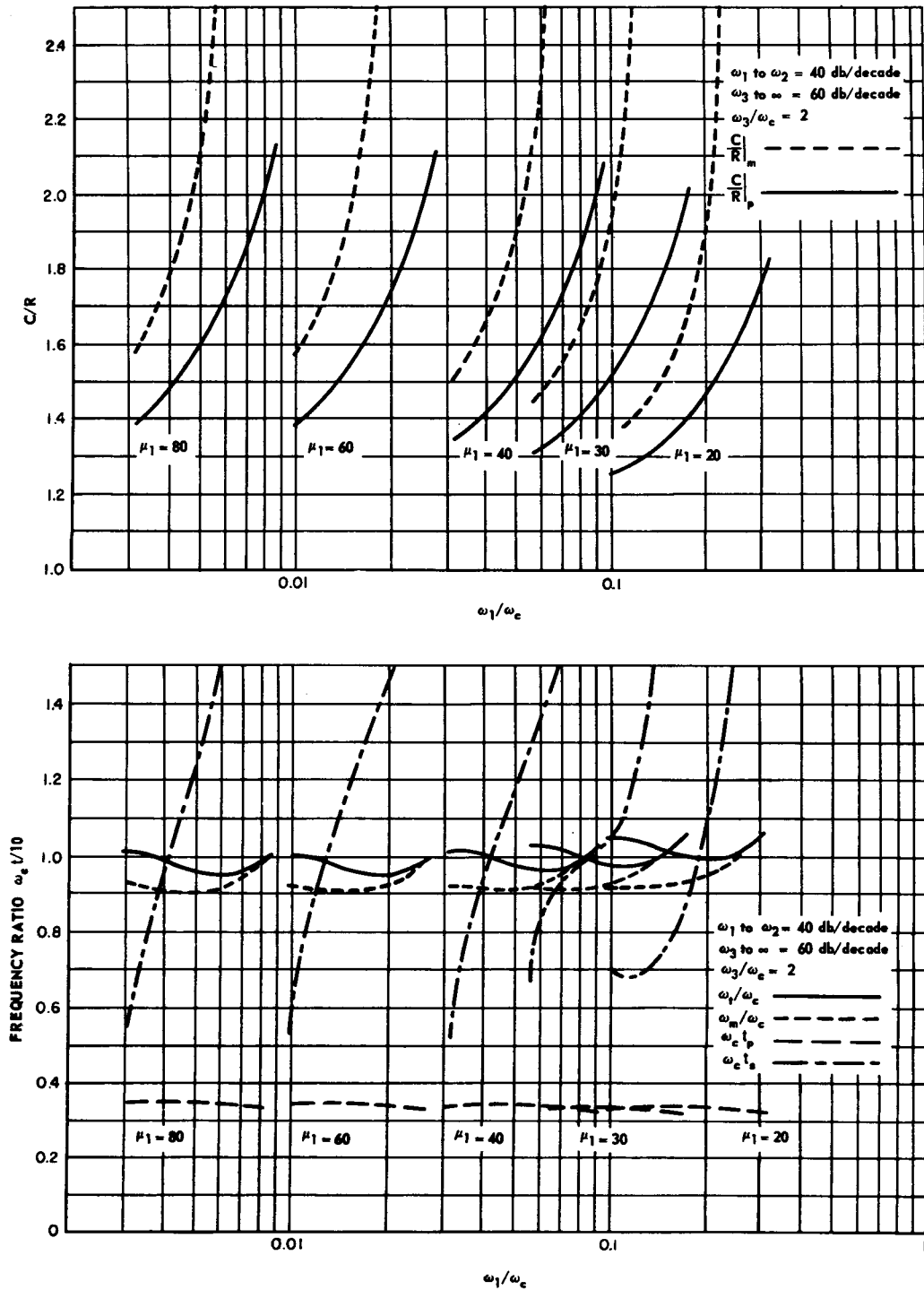


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 12 of 18)

THEORY

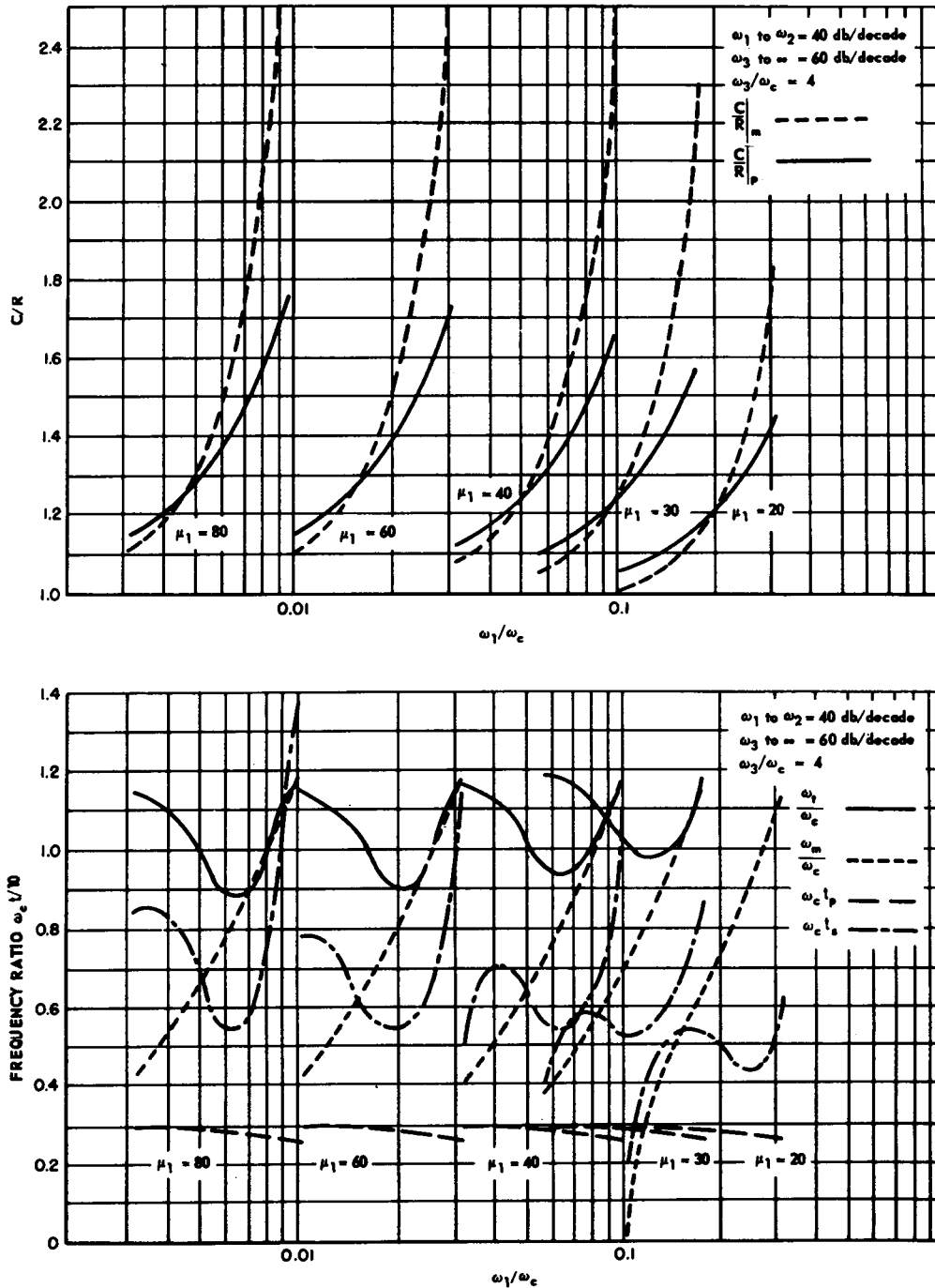


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 13 of 18)

PERFORMANCE EVALUATION

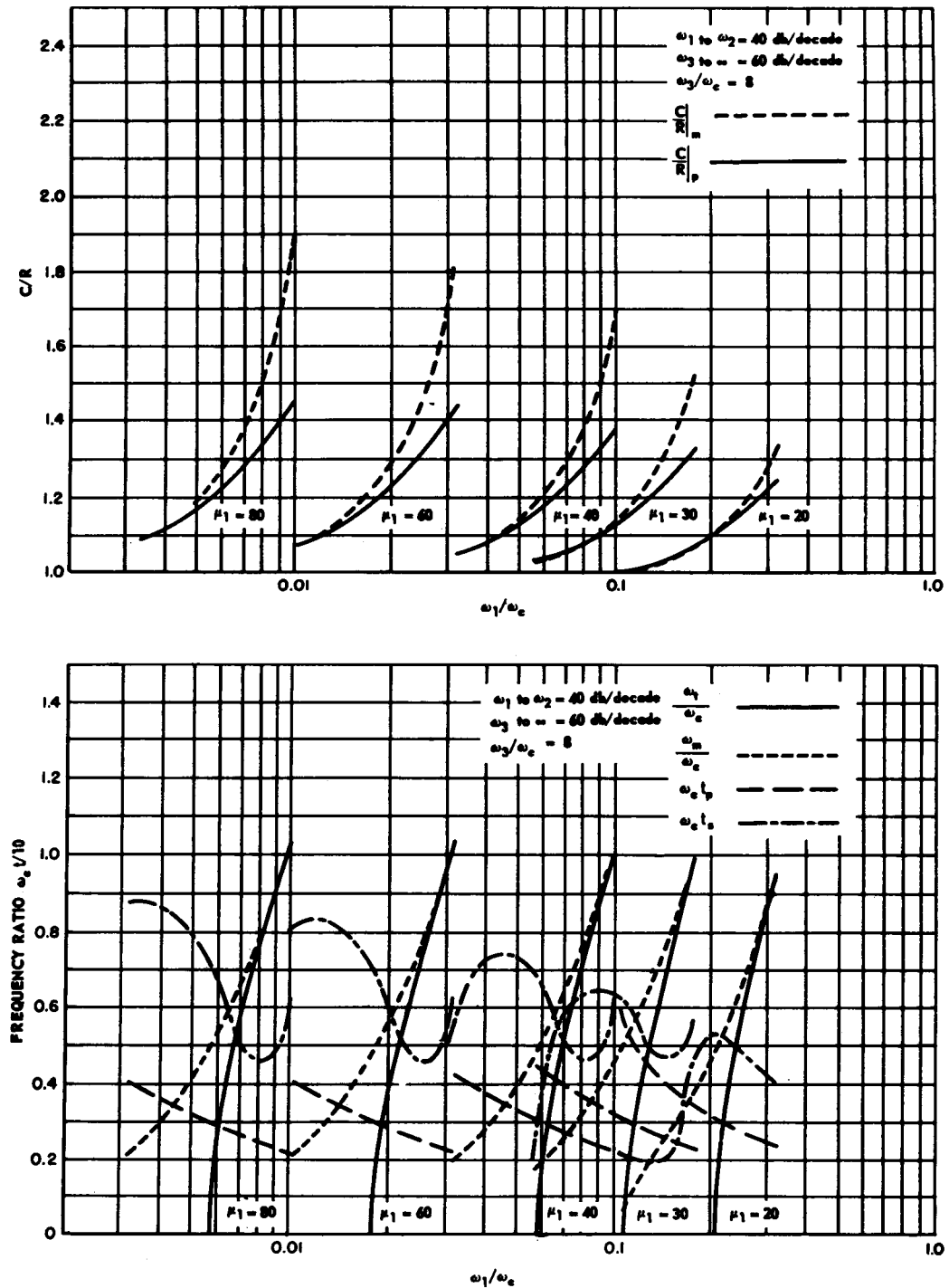


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 14 of 18)

THEORY

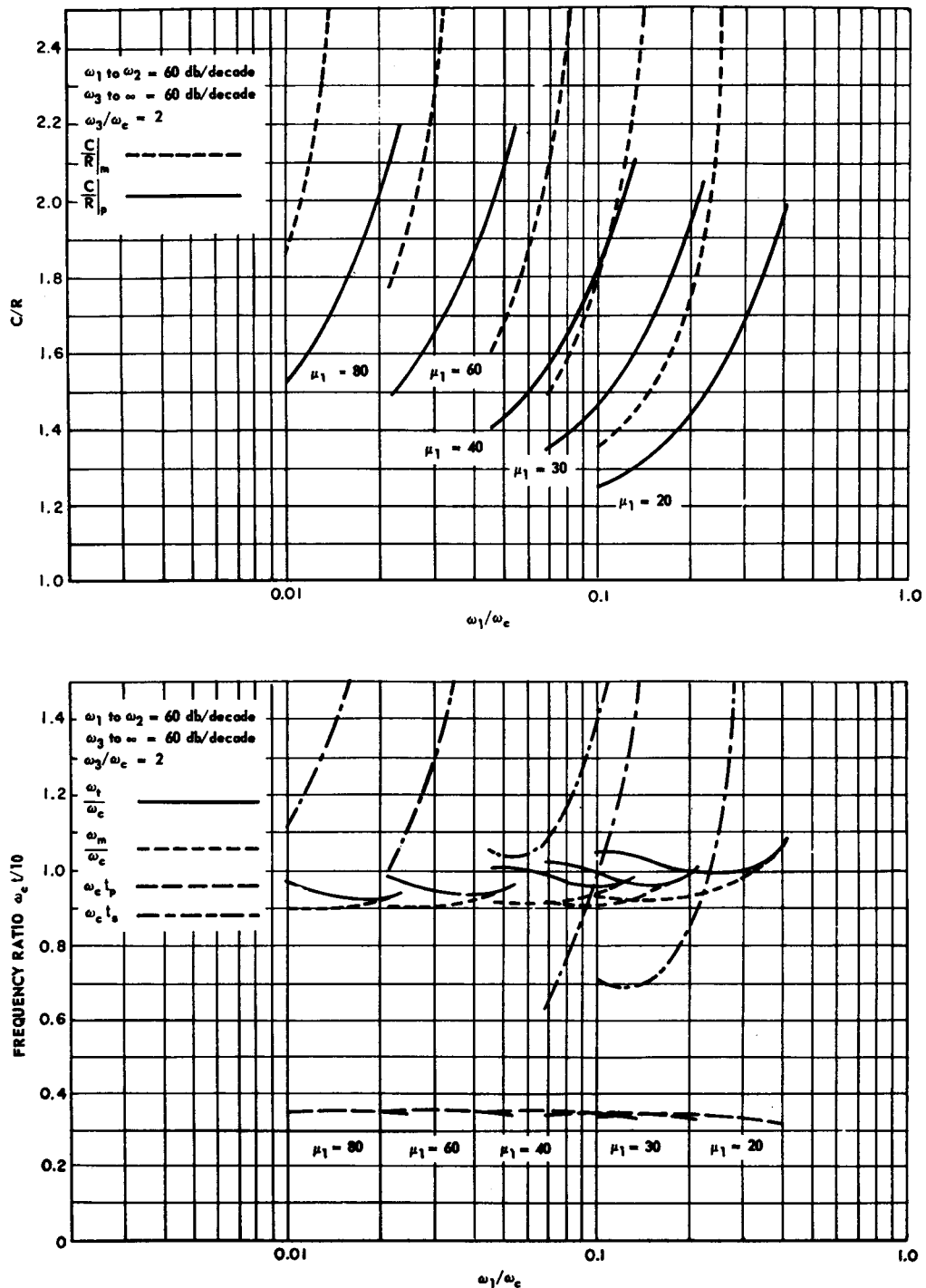


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 15 of 18)

PERFORMANCE EVALUATION

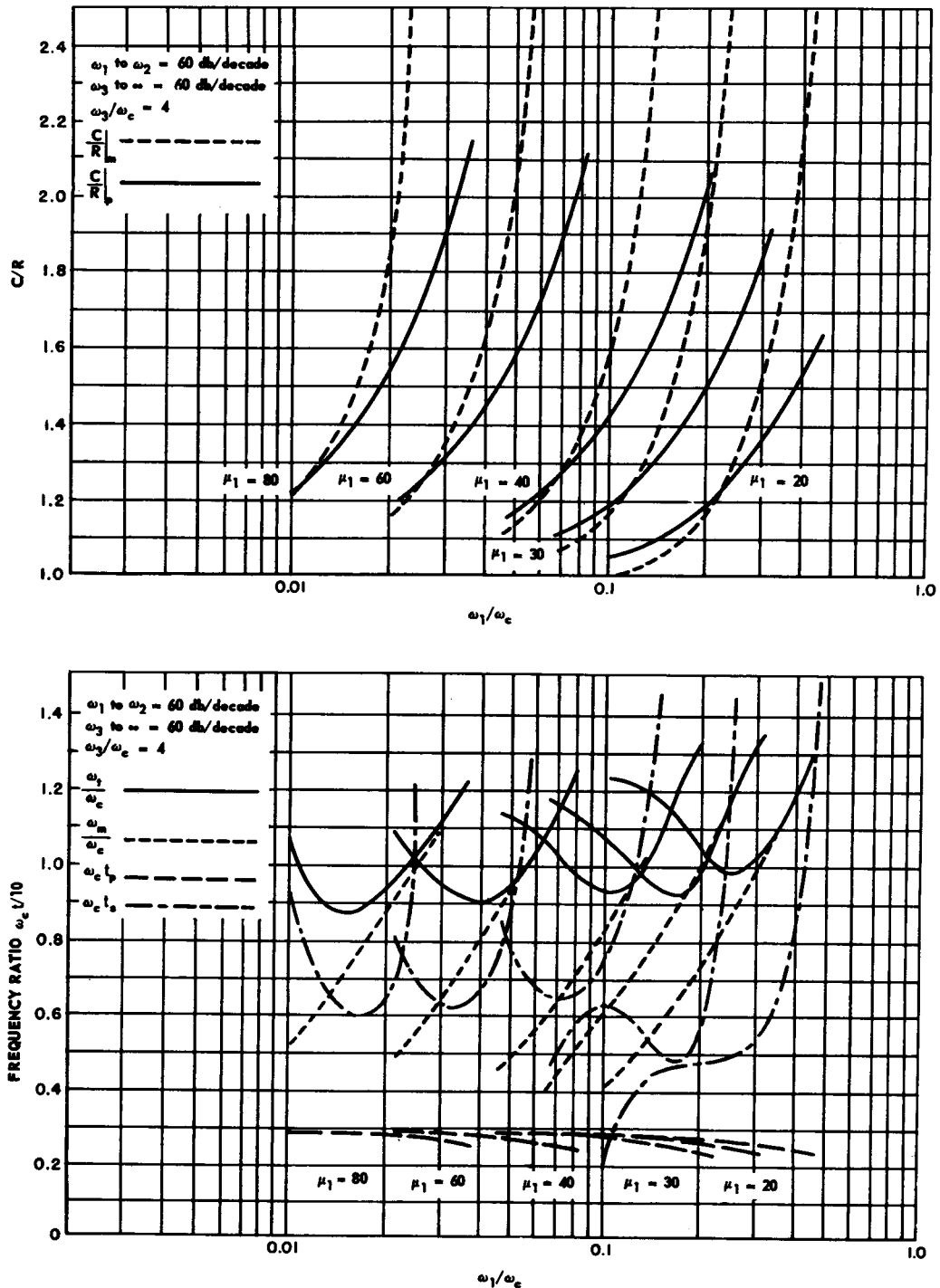


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 16 of 18)

THEORY

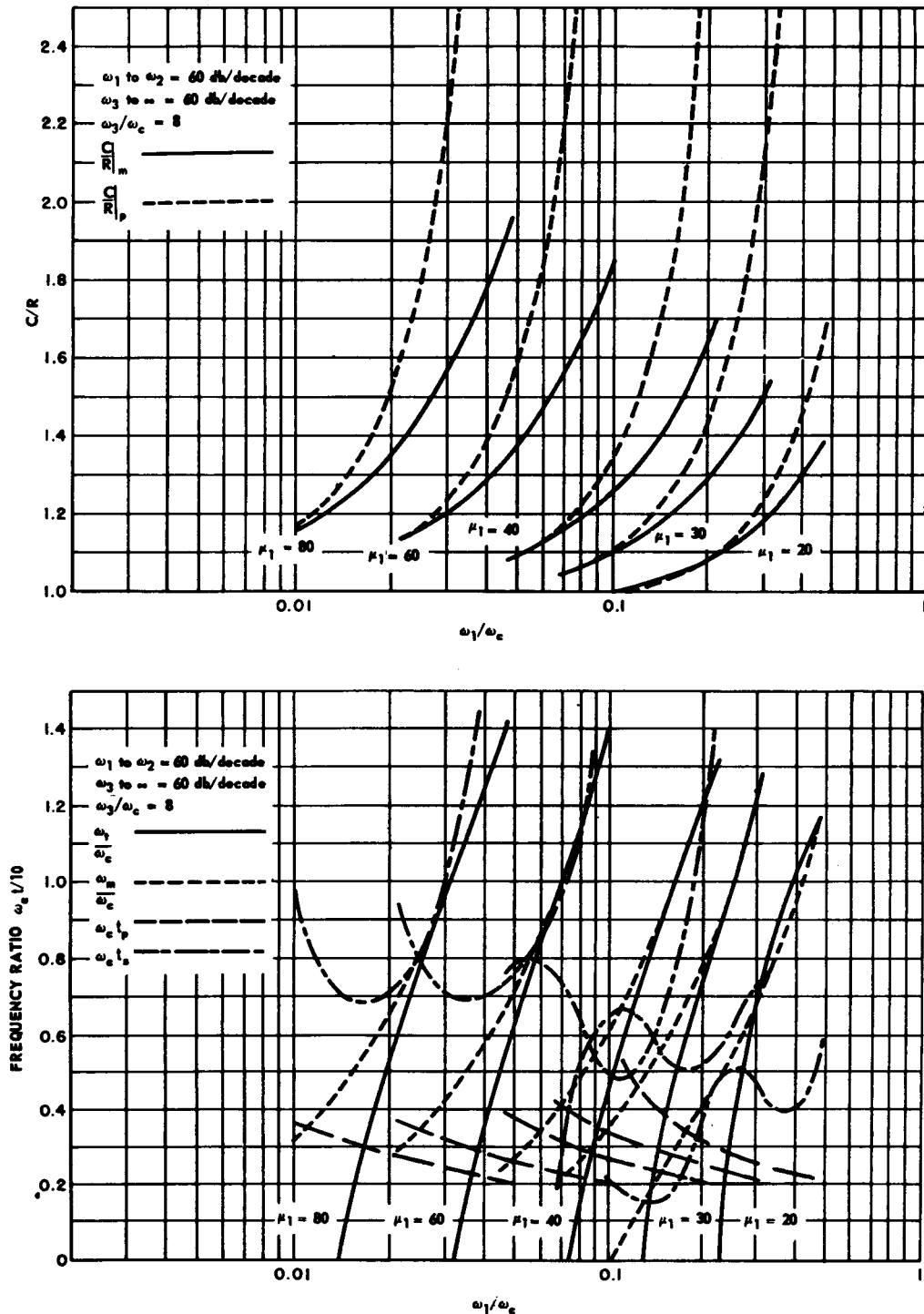


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 17 of 18)

PERFORMANCE EVALUATION

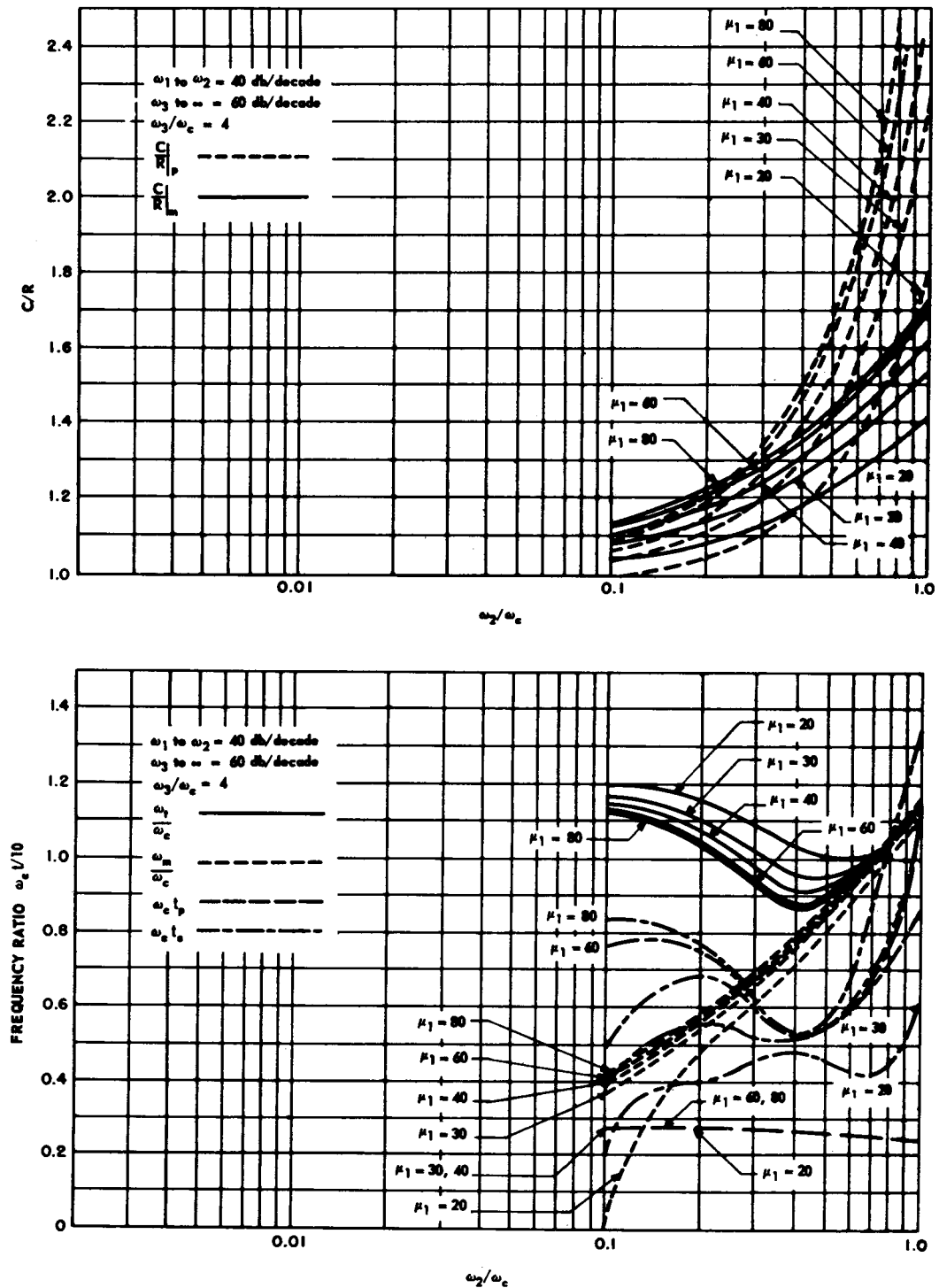


Fig. 7-18 Comparison of steady-state frequency response characteristics and transient response following a step function of input as a function of ω_1/ω_c . (Sheet 18 of 18)

THEORY

Inspection of Fig. 7-19 and Eq. (7-48) shows that the low-frequency zeros of $G(s)$ occur at $-\omega_2$ and $-\omega_3$ and that the high-frequency poles of $G(s)$ consist of a double pole at $-\omega_4$. Therefore, the poles of the first approximation to $C(s)/R(s)$ occur at $-\omega_2$ (single-order), $-\omega_3$ (single-order), and $-\omega_4$ (double-order), plus one pole at $-\omega_c$. The zeros of $C(s)/R(s)$ are the zeros of $G(s)$, $-\omega_2$ and $-\omega_3$. As a result, the first approximation to $C(s)/R(s)$ is

$$\begin{aligned}\frac{C(s)}{R(s)} &\cong \frac{\omega_c \omega_4^2}{(s + \omega_c)(s + \omega_4)^2} \\ &\cong \frac{K_o}{(s + \omega_c)} \frac{(\text{zeros of } G(s) \text{ above } \omega_c)}{(\text{poles of } G(s) \text{ above } \omega_c)}\end{aligned}\quad (7-49)$$

The first approximation to $E(s)$ is found to be

$$\begin{aligned}\frac{E(s)}{R(s)} &\cong \frac{s(s + \omega_1)^2}{(s + \omega_c)(s + \omega_2)(s + \omega_3)} \\ &\cong \frac{1}{(s + \omega_c)} \frac{(\text{poles of } G(s) \text{ below } \omega_c)}{(\text{zeros of } G(s) \text{ below } \omega_c)}\end{aligned}\quad (7-50)$$

The approximate factors of $1 + G(s)$ are given by

$$1 + G(s) \cong \frac{(s + 0.2)(s + 1)(s + 4)}{(s + 16)^2} \quad (7-51)$$

To evaluate the shifts from the approximate factors given in Eq. (7-51), Eqs. (7-46) and

(7-47) are used. The numerical form of $G(s)$ is

$$G(s) = \frac{4 \times 16^2 (s + 0.2)(s + 1)}{s(s + 0.04)^2 (s + 16)^2} \quad (7-52)$$

Since $\omega_c = 4$, the numerical method should be tried for the factors $(s + 0.2)$, $(s + 1)$, and $(s + 16)$. The shift from the approximate pole at $s = -0.2$ to the true pole is given by

$$\delta_2 \cong - \left. \frac{(s + 0.2)}{G(s)} \right|_{s=-0.2} = 0.0016 \quad (7-53)$$

Since this quantity is small, the true pole lies at

$$s = -0.2 + 0.0016 = -0.1984$$

For the approximate pole at $s = -\omega_3 = -1$, the shift to the true pole is given by

$$\delta_3 \cong - \left. \frac{(s + 1)}{G(s)} \right|_{s=-1} = -0.253 \quad (7-54)$$

This quantity is reasonably large so that a second approximation to δ_3 is made by evaluating the right side of Eq. (7-54) at $s = (-1) + (-0.253)$, instead of at $s = -1$, yielding

$$\delta_3 \cong - \left. \frac{(s + 1)}{G(s)} \right|_{s=-1.253} = -0.373 \quad (7-55)$$

The shift is still not too well approximated since the change from Eq. (7-54) to (7-55) is significant. The third approximation to δ_3 is obtained by evaluating Eq. (7-55) at $s = (-1) + (-0.373)$, instead of at $s = -1.253$, yielding

$$\delta_3 \cong - \left. \frac{(s + 1)}{G(s)} \right|_{s=-1.373} = -0.435 \quad (7-56)$$

Since the original approximate closed-loop pole at $s = -1$ was fairly close to the cross-over frequency $\omega_c = 4$, it is to be expected that the process of determining the shift δ_3 will converge fairly slowly. The succeeding approximations to the value of the pole are as follows:

-1.468 , -1.485 , -1.496 , -1.503 , -1.507 , -1.509 , and finally -1.510 .

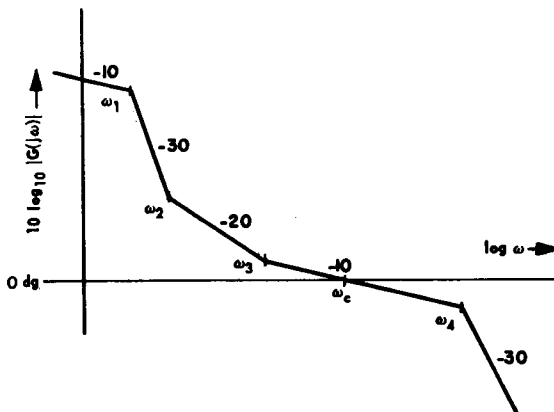


Fig. 7-19 Sketch of open-loop asymptote function.

For the approximate double-order pole at $s = -16$, the shift is given by

$$(\delta_4)^2 \cong -(s + 16)^2 G(s) \Big|_{s=-16} = 59.6 \quad (7-57)$$

or

$$\delta_4 = \pm 7.7$$

The original double-order pole splits in two, one pole moving toward ω_c and the other pole moving away from ω_c . Only the pole moving away from ω_c can be determined by the numerical method. The second approximation to the shift of this pole is $s = (-16) + (-7.7) = -23.7$. The third approximation to the shift of this pole is obtained by evaluating Eq. (7-57) at $s = -23.7$, instead of at $s = -16$, yielding

$$(\delta_4)^2 \cong -(s + 16)^2 G(s) \Big|_{s=-23.7} = \pm 6.4 \quad (7-58)$$

or $s = -22.4$ is the third approximation to the pole. Keeping track of the negative shift in this case, the succeeding values of the pole are found to be -22.6 and finally -22.56 .

Thus, three of the five "exact" factors of $1 + G(s)$ are $(s + 0.1984)$, $(s + 1.51)$, and $(s + 22.56)$. Dividing these factors out yields the remaining complex poles at $s = -4.35 \pm j3.3$. The "exact" close-loop response C/R is therefore

$$\frac{C(s)}{R(s)} = \frac{1014(s + 0.2)(s + 1)}{(s + 0.1984)(s + 1.51)(s + 22.56)(s^2 + 8.69s + 30)} \quad (7-59)$$

Although the detail with which this example has been presented may make the procedure seem laborious, actually it is extremely rapid even when the rate of convergence of the successive approximations is relatively slow. Note, also, that any desirable degree of accuracy can be maintained.

In other papers ^(23,25), Biernson gives a very good summary of approximate relations between the open-loop frequency response and the closed-loop transient response. If ω_{cm} is defined as the frequency at which the magnitude of $G(j\omega)$ crosses the 0-dB line (magnitude crossover frequency), then

(a) The *rise time* is approximately $1/\omega_{cm}$, where the rise time is defined as the time for the output response to a unit-step input to reach 0.63.

(b) The *peak error* for a unit-ramp input is approximately $1/\omega_{cm}$.

(c) The *peak output response* for a unit-impulse input is approximately ω_{cm} .

(d) The *peak overshoot* in the output response to a unit-step input is best determined from M_p by means of Figs. 7-14 and 7-15.

(e) The *settling time* t_s is approximated by the settling time of the equivalent second-order system unless $G(s)$ has low-frequency zeros produced by integral networks.

(f) If a first-order lag network (integral network) has been used to compensate the system, the peak overshoot of the output response to a unit-step input will be increased. If T_c is the time constant of the lag network, an additional transient term Ae^{-t/T_c} is added to the step response, where $A \cong 1/T_c\omega_c$ and ω_c is the *asymptote* crossover frequency.

(g) If an integral network is added to a system, the rate of decay of the error response to a unit-ramp input response is determined by the time constant T_c of the integral network. This response can be sketched from the following considerations:

(1) The response initially rises at the same rate as the input.

(2) The peak of the response is $1/\omega_{cm}$.

(3) If ω_c is the asymptote crossover frequency, the tail of the response is approximately the tail of an exponential with time constant T_c starting from $1/\omega_c$ at $t=0$ and falling to $1/K_v$ at $t = \infty$, where K_v is the velocity constant.

(h) The *maximum* time delay by which the output response to a unit-ramp input lags the input ramp is approximately equal to the rise time (0.63-value) of the output response to a unit-step input.

7-2 ERROR COEFFICIENTS^(17,23,27,31,32)

Paragraph 7-1 discusses the performance of feedback-control systems in terms of transient response and frequency response. Both of these views are intimately connected and stem from the impulse-response and convolution-integral description which forms the basis for all performance-evaluation methods. Unfortunately, it is generally true that the evaluation of performance by any of the foregoing methods is very laborious when the input is an arbitrary (but definable) time function. The error-coefficient method is a technique which aids the designer in such a case. Paragraph 3-6 shows that the error response of a system to a specified input can be expressed in terms of the input, its derivatives, and a set of *error coefficients* derivable from the transfer function of the system. The expression for the error response is

$$e(t) = e_0 r(t) + e_1 r'(t) + e_2 r''(t) + \dots \quad (7-60)$$

where the error coefficients e_0, e_1, e_2, \dots are the coefficients of the Maclaurin series expansion of the error-to-input transfer function $E(s)/R(s)$, i.e.,

$$\frac{E(s)}{R(s)} = e_0 + e_1 s + e_2 s^2 + \dots \quad (7-61)$$

This expansion is valid everywhere except where the input or any of its derivatives are discontinuous. For practical purposes, only a few terms of the expansion are used to evaluate the error response. However, the expansion cannot be used near points of discontinuity of r, r', r'', \dots if accurate results are sought. Thus, for example, if a step discontinuity occurs in the input $r(t)$, the expansion is invalid for a time interval extending from the instant t_0 at which the step occurs to the time $(t_0 + t_s)$, where t_s is the settling time of the transient error response to the step (time for the error transient to fall to 2% of its initial value). Obviously, the step can be ignored if it is small compared to the remaining terms of the expansion in the interval $t_0 < t < (t_0 + t_s)$.

Biernson⁽²³⁾ has suggested that the foregoing difficulty can be resolved by examining r, r', r'', \dots for discontinuities and subtracting these discontinuities from the corresponding functions. The remaining functions will all be continuous, and the expansion can be applied over the entire time range of interest. Then, the effects of the discontinuities in r, r', r'', \dots are added to the response. In this procedure, a discontinuity in a function is considered to occur if the function rises (or decays) more abruptly than the corresponding transient response to the discontinuity. In comparing the rise rate of the two curves, a convenient criterion is to compare the times for the two curves to reach 63% of the initial rise (or decay) of the curves.

A convenient procedure for determining the error coefficients required to satisfy performance specifications is the following:

(a) Given the input $r(t)$ and the maximum allowable error e_{max} which can be tolerated at any time, determine the derivatives r', r'', r''', \dots of the input $r(t)$.

(b) Assume values of the error coefficients so that the maximum value of each component term in the expansion [Eq. (7-60)] is equal to e_{max} .

(c) Add the curves obtained in (b) to obtain the first trial value of the error response $e_1(t)$.

(d) There will be times in which the first trial $e_1(t)$ will exceed e_{max} . Referring to the curves found in (b), decide which of the functions r, r', r'' have their maxima in regions where $e_1(t)$ exceeds e_{max} .

(e) Reduce the assumed values of the error coefficients in (b) associated with the functions found in (d).

(f) Add the adjusted curves found in (e) to those functions [found in (b)] which have not been changed. Determine whether e_{max} is now exceeded and, if so, repeat (d) and (e).

PERFORMANCE EVALUATION

Once the error coefficients have been specified by means of a procedure similar to the foregoing, the design of the system can be carried out as follows:

(a) The system is designed to meet all other specifications on transient response and frequency response such as bandwidth, M_p , P_{ov} , etc.

(b) The error coefficients of the system designed in (a) are found in terms of the system parameters and the gain.

(c) If any of the error coefficients found in (b) exceed their specified values, they may be reduced by increasing the system gain, if possible, by the introduction of low-frequency dipole *lead* functions in the pole-zero configuration of the *open-loop* transfer function $G(s)$, or by the feedback-compensation technique suggested by King.⁽¹⁷⁾

(d) The specifications on transient re-

sponse and frequency response are rechecked, and step (c) is modified if necessary.

A warning should be added. Whenever attempts are made to reduce one or more error coefficients of a system by the methods suggested above, it is possible that higher-order error coefficients may increase. Therefore, if by the addition of a low-frequency lead dipole, an error coefficient can be reduced to zero, a check should be made to insure that higher coefficients have not been increased excessively. In addition, it is generally true that low-frequency poles in a transfer function tend to increase the settling time of the response of the system to steps in the higher input-derivative functions. Therefore, if the actual input being examined has discontinuities in one or more of its higher derivatives, the effect of the longer settling time in the response to these discontinuities must be determined.

7-3 PERFORMANCE INDICES^(1,2,9,14,16,18)

A *performance index* P is a single number which is used as an *indirect* measure of system performance. Other measures of system performance have already been considered, such as the various commonly used parameters M_p , bandwidth, rise time, peak overshoot, etc. However, these parameters provide only a partial description of performance since, in a sense, only part of the corresponding response is described by each. To be sure, if enough of these "*response parameters*" (for want of a better term) are known, an accurate description of the corresponding response is possible. That is, the "*response parameters*" may be considered *direct* descriptions of the shape of their associated responses. However, since a response function is continuous, theoretically an infinite number of response parameters are necessary to describe the response. To get around the use of a multitude of response parameters, a

performance index may be used. The use of performance indices is an attempt to replace the functional description of the performance of a system through its response parameters by a numerical description that rates the system performance with a single number.

Paragraphs 8-1 and 8-2 describe various techniques for using performance indices. This section merely presents the commonly used indices together with the input conditions for which they apply. Table 7-1 is a summary of these indices. In practice, the performance index corresponding to the specified input is found, and the system is adjusted to optimize (minimize or maximize) the index. The indices P_1 , P_3 , P_5 , P_6 , P_7 , and P_8 can be used in purely analytical procedures. However, P_2 and P_4 are not treated analytically but rather through the use of analog computers.

THEORY

TABLE 7-1 COMMON PERFORMANCE INDICES

| Index | Input | Description or Name | Reference |
|-----------------------------------|------------|--|------------|
| $P_1 = \int_0^\infty e \, dt$ | Transient | Control area | 1,2 * |
| $P_2 = \int_0^\infty e \, dt$ | Transient | Integral absolute error (IAE) | 9 |
| $P_3 = \int_0^\infty t e \, dt$ | Transient | | 1 |
| $P_4 = \int_0^\infty t e \, dt$ | Transient | Integral-time-multiplied absolute error (ITAE) | 14,16,18 |
| $P_5 = \int_0^\infty e^2 \, dt$ | Transient | Integral-square error (ISE) | 9,18,33,34 |
| $P_6 = \int_0^\infty t e^2 \, dt$ | Transient | | |
| $P_7 = P_3/P_1^2$ | Transient | | 1 |
| $P_8 = \overline{e^2}$ | Stochastic | Mean-square error (MSE) | 34,35 |

* BIBLIOGRAPHY

- 1 P. T. Nims, "Some Design Criteria for Automatic Controls", *Trans. AIEE*, Vol. 70, Part I, pp. #606-611, 1951.
- 2 T. M. Stout, "A Note on Control Area", *J. Appl. Phys.*, Vol. 21, pp. #1129-1131, November, 1950.
- 3 M. R. Aaron, "Synthesis of Feedback Control Systems by Means of Pole and Zero Location of the Closed-Loop Function", *Trans. AIEE*, Vol. 70, Part II, pp. #1439-1445, 1951.
- 4 D. Herr and I. Gerst, "The Analysis and an Optimum Synthesis of Linear Servomechanisms", *Trans. AIEE*, Vol. 66, pp. #959-970, 1947.
- 5 J. H. Mulligan, Jr., "The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems", *Proc. IRE*, Vol. 37, pp. #516-529, May, 1949.
- 6 D. W. Russel and C. H. Weaver, "Synthesis of Closed-Loop Systems Using Curvilinear Squares to Predict Root Location", *Trans. AIEE*, Vol. 71, Part II, pp. #95-104, 1952.
- 7 H. Harris, Jr., M. J. Kirby, and E. F. Von Arx, "Servomechanism Transient Performance from Decibel-Log Frequency Plots", *Trans. AIEE*, Vol. 70, Part II, pp. #1452-1459, 1951.
- 8 H. Chestnut and R. W. Mayer, "Comparison of Steady-State and Transient Performance of Servomechanisms", *Trans. AIEE*, Vol. 68, Part I, pp. #765-777, 1949.
- 9 F. C. Fickeison and T. M. Stout, "Analogue Methods for Optimum Servomechanism Design", *Trans. AIEE*, Vol. 71, Part II, pp. #244-250, 1952.

PERFORMANCE EVALUATION

- 10 M. J. Kirby and D. C. Beaumariage, "Relative Stability of Closed-Loop Systems", *Trans. AIEE*, Vol. 72, Part II, pp. #22-43, 1953.
- 11 G. Biernson, "Quick Methods for Evaluating the Closed-Loop Poles of Feedback Control Systems", *Trans. AIEE*, Vol. 72, Part II, pp. #53-70, 1953.
- 12 N. L. Kusters and W. J. M. Moore, "A Generalization of the Frequency Response Method for Study of Feedback Control Systems", in *Automatic and Manual Control*, edited by A. Tustin, Butterworth Scientific Publications, London, England, 1952.
- 13 Y. Chu, "Correlation Between Frequency and Transient Responses of Feedback Control Systems", *Trans. AIEE*, Vol. 72, Part II, pp. #81-92, 1953.
- 14 D. Graham and R. C. Lathrop, "The Synthesis of 'Optimum' Transient Response: Criteria and Standard Forms", *Trans. AIEE*, Vol. 72, Part II, pp. #273-288, 1953.
- 15 J. G. Linvill, "The Approximation with Rational Functions of Prescribed Magnitude and Phase Characteristics", *Proc. IRE*, Vol. 40, pp. #711-721, 1952.
- 16 R. C. Lathrop and D. Graham, "The Transient Performance of Servomechanisms with Derivative and Integral Control", *Trans. AIEE*, Vol. 73, Part II, pp. #10-17, 1954.
- 17 L. H. King, "Reduction of Forced Error in Closed-Loop Systems", *Proc. IRE*, Vol. 41, pp. #1037-1042, August, 1953.
- 18 D. Graham and R. C. Lathrop, "The Influence of Time Scale and Gain on Criteria for Servomechanism Performance", *Trans. AIEE*, Vol. 73, Part II, pp. #153-158, 1954.
- 19 A. H. Zemanian, "Bounds Existing on the Time and Frequency Responses of Various Types of Networks", *Proc. IRE*, Vol. 42, pp. #835-839, May, 1954.
- 20 A. H. Zemanian, "Further Effects of the Pole and Zero Locations on the Step Response of Fixed, Linear Systems", *Trans. AIEE*, Vol. 74, Part II, pp. #52-55, 1955.
- 21 H. Thal-Larsen, "Frequency Response from Experimental Nonoscillatory Transient-Response Data", *Trans. AIEE*, Vol. 74, Part II, pp. #109-113, 1955.
- 22 D. W. St. Clair, "Step Response as a Short Cut to Frequency Response", *Proc. ISA*, Vol. 7, pp. #96-101, 1952.
- 23 G. A. Biernson, "A Simple Method for Calculating the Time Response of a System to an Arbitrary Input", *Trans. AIEE*, Vol. 74, Part II, pp. #227-245, 1955.
- 24 M. E. Clynes, "Simple Analytic Method for Linear Feedback System Dynamics", *Trans. AIEE*, Vol. 74, Part II, pp. #377-382, 1955.
- 25 G. A. Biernson, "Estimating Transient Response from Open-Loop Frequency Response", *Trans. AIEE*, Vol. 74, Part II, pp. #388-402, 1955.
- 26 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. 1, pp. #398-439, John Wiley & Sons, Inc., New York, N. Y., 1951.
- 27 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. 2, pp. #1-35, 43-65, John Wiley & Sons, Inc., New York, N. Y., 1955.
- 28 J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. #34-87, 278-296, 344-390, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.
- 29 R. C. Seamans, Jr., B. P. Blasingame, and G. C. Clementson, "The Pulse Method for the Determination of Aircraft Performance", *J. Aeronaut. Sci.*, Vol. 17, pp. #22-38, January, 1950.

THEORY

- 30 E. A. Guillemin, "Computational Techniques which Simplify the Correlation between Steady-State and Transient Responses of Filters and Other Networks", *Proc. Nat. Electronics Conf. 1953*, Vol. 9, 1954.
- 31 J. L. Bower, "A Note on the Error Coefficients of a Servomechanism", *J. Appl. Phys.*, Vol. 21, p. #723, July, 1950.
- 32 E. Arthurs and L. Martin, "A Closed Expansion of the Convolution Integral (A Generalization of Servomechanism Error Coefficients)", *J. Appl. Phys.*, Vol. 26, pp. #58-60, January, 1955.
- 33 A. C. Hall, *The Analysis and Synthesis of Linear Servomechanisms*, The Technology Press, Massachusetts Institute of Technology, Cambridge, Mass., 1943.
- 34 G. C. Newton, Jr., "Compensation of Feedback-Control Systems Subject to Saturation", *J. Franklin Inst.*, Vol. 254, pp. #281-296, 391-413, 1952.
- 35 J. H. Laning, Jr., and R. H. Battin, *Random Processes in Automatic Control*, McGraw-Hill Book Co., Inc., New York, N. Y., 1956.
- 36 G. S. Brown and D. P. Campbell, *Principles of Servomechanisms*, John Wiley & Sons, Inc., New York, N. Y., 1948.
- 37 J. G. Linvill, "The Selection of Network Functions to Approximate Prescribed Frequency Characteristics", *MIT Research Laboratory of Electronics Tech Rept. 145*, March 14, 1950.
- 38 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. 1, pp. #401 and 417-434, John Wiley & Sons, Inc., New York, N. Y., 1951.

CHAPTER 8

OPTIMIZATION METHODS FOR TRANSIENT AND STOCHASTIC INPUTS***8-1 CRITERIA OF PERFORMANCE**

In Par. 7-3 it was indicated that the conventional measures of performance such as rise time, peak overshoot, solution time, M_p , etc. were merely partial descriptions of the frequency response of a system or the shape of a particular transient response. As a result, an adequate description of system behavior requires a fair number of response or performance parameters. To avoid using a multiplicity of response parameters (numbers which describe the response such as M_p , peak overshoot, bandwidth, rise time, etc.) attempts have been made to describe system behavior in terms of performance indices. A performance index is a single number which can be used as a criterion of performance. The pertinent performance indices used are those directly related to system error since *error* is the basic determinant of the "goodness" of a system. The most common performance indices are those listed in Par. 7-3.

When performance indices are used in system design, the usual procedure is to minimize the index if it is a direct measure of error. With a given index one also associates the specified input to the system. Several approaches can be used in carrying out the minimization procedure.

In one approach it may be assumed that all but a few of the system parameters are specified. Then, the optimization procedure involves the adjustment of the *free* parameters so as to minimize the performance index.

Such a procedure is called a *fixed-configuration minimization* method or technique since the form of the system is specified and only the numerical values of the free parameters are sought.

In another approach nothing is assumed about the configuration of the system. Here the entire impulse response of the system is varied to minimize the performance index. This procedure is called a *free-configuration minimization method or technique*.

Of the two procedures, the easier one to apply and the one more commonly used is the fixed-configuration technique since the process of minimization can be carried out by differentiating the performance index with respect to the free parameters and setting the resulting partial derivatives equal to zero. The fixed-configuration technique is also easy to apply when use is made of an analog computer. The free-configuration method, on the other hand, is less commonly used because it can only be applied by the use of the calculus of variations since in this case a system *function* (rather than a system parameter) is varied to obtain a minimum.

In practice, the application of optimization methods can lead to failure when one is not cognizant of the limitations of the mathematical model that represents the physical system. The optimum system often requires cancellation of the characteristics of the fixed elements of the system, resulting in an unnecessarily wide-band performance and concurrent nonlinear operation. To avoid this, constraints may be placed on signal levels or on

*By L. A. Gould

bandwidth. Constrained optimization brings the designer closer to the practical limitations of the system and serves to guide system design in a realistic way.

The primary advantages of optimization procedures (as contrasted with conventional trial-and-error procedures) are twofold.

First, the designer is able, through optimization, to decide whether a given set of specifications is compatible. Second, the designer can decide whether a compatible set of specifications can be satisfied when bounded by constraints.

8-2 OPTIMUM SYNTHESIS OF FIXED-CONFIGURATION SYSTEMS

8-2.1 TRANSIENT INPUTS

For transient inputs the integral-square error criterion (ISE) is commonly used^(6,7,10) to obtain optimum synthesis of a fixed-configuration system. If discussion is limited to unity-feedback systems, where the desired output is the input, then the actuating signal $e(t)$ is equal to the system error $y_e(t)$. The ISE criterion is then

$$I_y = \int_{-\infty}^{+\infty} y_e^2(t) dt \quad (8-1)$$

The evaluation of the integral in Eq. (8-1) is facilitated by the application of Parseval's theorem:

$$\int_{-\infty}^{+\infty} x^2(t) dt = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} X(s) X(-s) ds \quad (8-2)$$

where $X(s)$ is the Fourier transform of $x(t)$. Thus, Eq. (8-1) can be written

$$I_y = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} Y_e(s) Y_e(-s) ds \quad (8-3)$$

The procedure for minimizing the ISE is as follows:

(a) Express the Fourier transform of the error as a function of the complex frequency s . This function will involve the free parameters of the system as unknown coefficients.

(b) Express I_y in terms of $Y_e(s)$ by means of Eq. (8-3). If $Y_e(s)$ is rational, the form of I_y will be

$$I_y = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} \frac{C(s) ds}{D(s) D(-s)} \quad (8-4)$$

where $C(s)$ and $D(s)$ are polynomials in s .

(c) Evaluate the integral in Eq. (8-4). Definite integrals of this form have been evaluated in terms of the coefficients of the polynomials in the integrand.^(8,16) A brief table of such integrals is presented in Table 8-1 where the evaluation has been carried out for $s = j\omega$. At this point, I_y is expressed as a function of the free parameters p_1 through p_k , i.e.,

$$I_y = I_y(p_1, p_2, \dots, p_k) \quad (8-5)$$

(d) Adjust the free parameters p_1, p_2, \dots so as to minimize I_y . This can be accomplished analytically by solving the k simultaneous equations

$$\frac{\partial I_y}{\partial p_i} = 0 \quad (i = 1, 2, \dots, k) \quad (8-6)$$

However, it is often better to find the minimum graphically by working directly with I_y .

Example. A unity-feedback system has the fixed-element transfer function

$$G_f(s) = \frac{1}{s(T_f s + 1)(T_m s + 1)}$$

where

$$T_f = 0.01 \text{ second and}$$

$$T_m = 0.04 \text{ second}$$

OPTIMIZATION METHODS

The compensation $G_c(s)$ is a pure gain, i.e.,

$$G_c(s) = K_v, \text{ the velocity constant}$$

The input is a step of magnitude N_i , i.e.,

$$r(t) = N_{i0-1}(t)$$

The desired output is the input. The configuration that describes the problem appears in Fig. 8-1.

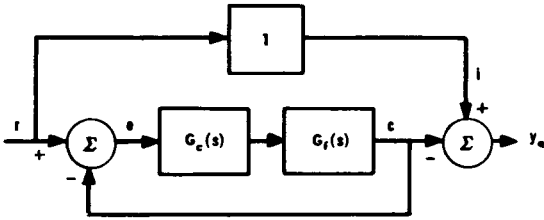


Fig. 8-1 Configuration for ISE minimization.

Solution. To find the value of K_v that minimizes the integral-square error I_y , we first find the error transform $Y_e(s)$. From the configuration in Fig. 8-1 it is evident that

$$Y_e(s) = E(s) = R(s) - C(s), \text{ and}$$

$$C(s) = \frac{G_c(s)G_f(s)}{1 + G_c(s)G_f(s)} R(s)$$

$$\text{Now, } R(s) = \frac{N_i}{s}.$$

So, by substituting the expressions given originally for $G_c(s)$ and $G_f(s)$ into the equation for $C(s)$ we find that the error transform is

$$Y_e(s) = N_i \frac{T_f T_m s^2 + (T_f + T_m)s + 1}{T_f T_m s^3 + (T_f + T_m)s^2 + s + K_v}$$

Substituting $Y_e(s)$ into Eq. (8-3) and letting $s = j\omega$, we find that

$$I_y = \frac{N_i^2}{2\pi} \int_{-\infty}^{+\infty} \frac{[c_4 \omega^4 + c_2 \omega^2 + c_0] d\omega}{[d_3 (j\omega)^3 + d_2 (j\omega)^2 + d_1 (j\omega) + d_0] [d_3 (-j\omega)^3 + d_2 (-j\omega)^2 + d_1 (-j\omega) + d_0]}$$

where

$$\begin{aligned} c_0 &= 1 & d_0 &= K_v \\ c_2 &= T_f^2 + T_m^2 & d_1 &= 1 \\ c_4 &= T_f^2 T_m^2 & d_2 &= T_f + T_m \\ & & d_3 &= T_f T_m \end{aligned}$$

Using I_3 of Table 8-1 to evaluate the integral above, we get

$$I_y = \frac{N_i^2}{2} \left\{ \frac{1 + \frac{T_f}{T_m} + \frac{T_m}{T_f} + \frac{1}{K_v} \left(\frac{1}{T_f} + \frac{1}{T_m} \right)}{\frac{1}{T_m} + \frac{1}{T_f} - K_v} \right\}$$

Numerically, this becomes

$$I_y = \frac{N_i^2}{2} \left[\frac{5.25 K_v + 125}{125 K_v - K_v^2} \right]$$

Inspection shows that $I_y \rightarrow 0$ if $K_v \rightarrow \infty$, but this solution is not allowed since the system would then be unstable. Differentiating I_y with respect to K_v and setting $dI_y/dK_v = 0$ yields

$$K_v^2 + 47.6 K_v - 2980 = 0$$

or

$$K_v = 35.8 \text{ or } -83.4$$

The negative value is not allowed so a velocity constant $K_v = 35.8 \text{ sec}^{-1}$ minimizes the integral-square error. The value of the minimum integral-square error is

$$I_{y_{min}} = 0.049 N_i^2$$

As a point of interest, for $K_v = 35.8 \text{ sec}^{-1}$, the value of the peak magnification is $10 \log_{10} M_p = 2.8 \text{ dg}$ which is a reasonable value.

Another optimization criterion is presented in a series of papers by Graham and Lathrop^(3,4,5) in which they have applied the integral-time-multiplied-absolute-error criterion (ITAE) to optimize the performance of

THEORY

standard-system forms. However, their procedure is limited to step inputs only. The ITAE criterion is

$$I_{ta} = \int_{-\infty}^{+\infty} t |y_e(t)| dt \quad (8-7)$$

Although the analytical application of the ITAE criterion is practically impossible, the performance index can be easily mechanized on an analog computer.

For systems exhibiting zero steady-state error for a step input (finite velocity constant), the standard form chosen was

$$\frac{C(s)}{R(s)} = \frac{1}{s^n + q_{n-1}s^{n-1} + \dots + q_1s + 1} \quad (8-8)$$

The denominator polynomials of the optimum systems (that minimize the ITAE for a step input) are listed in Table 8-2 for the first eight orders. Figure 8-2 shows the step responses of the optimum systems, and Fig. 8-3 shows the frequency responses of these systems.

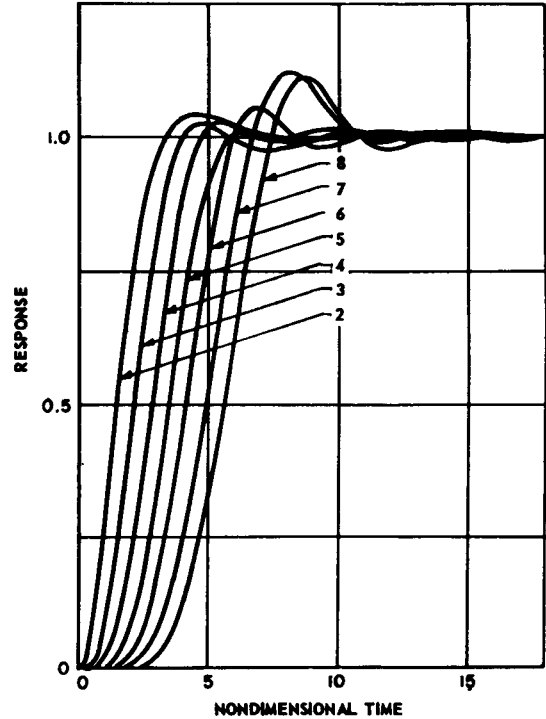


Fig. 8-2 Step-function responses of the optimum unit-numerator transfer systems, second to eighth orders.

By permission from *Transactions of the AIEE*, Volume 72, Part II, 1953, from article entitled "The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms", by Dunstan Graham and R. C. Lathrop.

TABLE 8-1 TABLE OF DEFINITE INTEGRALS

$$I_n = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{C_{2n-2}(\omega)}{D_n(j\omega) D_n(-j\omega)} d\omega$$

where

$$C_{2n-2}(\omega) = c_{2n-2}\omega^{2n-2} + c_{2n-4}\omega^{2n-4} + \dots + c_2\omega^2 + c_0$$

$$D_n(j\omega) = d_n(j\omega)^n + d_{n-1}(j\omega)^{n-1} + \dots + d_1j\omega + d_0$$

$$I_1 = \frac{c_0}{2d_1d_0}$$

$$I_2 = \frac{c_2d_0 + d_2c_0}{2d_2d_1d_0}$$

$$I_3 = \frac{c_4d_1d_0 + c_2d_3d_0 + c_0d_3d_2}{2d_3d_0(d_2d_1 - d_3d_0)}$$

$$I_4 = \frac{(d_2d_1 - d_3d_0)d_0c_6 + d_4d_1d_0c_4 + d_4d_3d_0c_2 + (d_3d_2 - d_4d_1)d_4c_0}{2d_4d_0(d_3d_2d_1 - d_4d_1^2 - d_3^2d_0)}$$

OPTIMIZATION METHODS

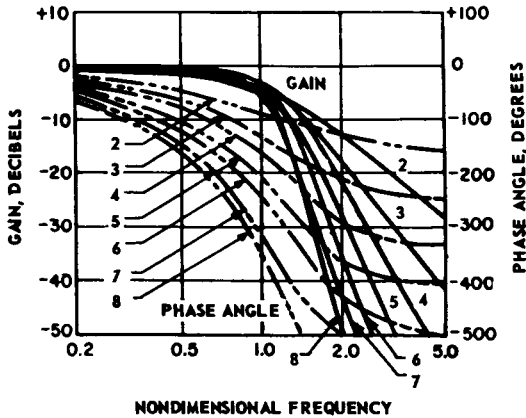


Fig. 8-3 Frequency responses of the optimum unit-numerator transfer systems.

By permission from *Transactions of the AIEE*, Volume 72, Part II, 1953, from article entitled "The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms", by Dunstan Graham and R. C. Lathrop.

For systems exhibiting zero steady-state error for a ramp input (infinite velocity constant, finite acceleration constant), the standard form chosen was

$$\frac{C(s)}{R(s)} = \frac{q_1 s + 1}{s^n + q_{n-1} s^{n-1} + \dots + q_2 s^2 + q_1 s + 1} \quad (8-9)$$

The denominator polynomials of the optimum systems of this type are listed in Table 8-3 for the second to sixth orders. The step responses of the optimum systems are shown in Fig. 8-4.

Additional matter such as optimum compensation of various systems and the effect of time scaling are also discussed by these authors^(4,5) with the use of the ITAE criterion as the performance index.

TABLE 8-2 THE MINIMUM ITAE STANDARD FORMS, ZERO-DISPLACEMENT-ERROR SYSTEMS

$$\begin{aligned} & s + \omega_0 \\ & s^2 + 1.4\omega_0 s + \omega_0^2 \\ & s^3 + 1.75\omega_0 s^2 + 2.15\omega_0^2 s + \omega_0^3 \\ & s^4 + 2.1\omega_0 s^3 + 3.4\omega_0^2 s^2 + 2.7\omega_0^3 s + \omega_0^4 \\ & s^5 + 2.8\omega_0 s^4 + 5.0\omega_0^2 s^3 + 5.5\omega_0^3 s^2 + 3.4\omega_0^4 s + \omega_0^5 \\ & s^6 + 3.25\omega_0 s^5 + 6.60\omega_0^2 s^4 + 8.60\omega_0^3 s^3 + 7.45\omega_0^4 s^2 + 3.95\omega_0^5 s + \omega_0^6 \\ & s^7 + 4.475\omega_0 s^6 + 10.42\omega_0^2 s^5 + 15.08\omega_0^3 s^4 + 15.54\omega_0^4 s^3 + 10.64\omega_0^5 s^2 + 4.58\omega_0^6 s + \omega_0^7 \\ & s^8 + 5.20\omega_0 s^7 + 12.80\omega_0^2 s^6 + 21.60\omega_0^3 s^5 + 25.75\omega_0^4 s^4 + 22.20\omega_0^5 s^3 + 13.30\omega_0^6 s^2 + 5.15\omega_0^7 s + \omega_0^8 \end{aligned}$$

By permission from *Transactions of the AIEE*, Volume 72, Part II, 1953, from article entitled "The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms", by Dunstan Graham and R. C. Lathrop.

TABLE 8-3 THE MINIMUM ITAE STANDARD FORMS, ZERO-VELOCITY-ERROR SYSTEMS

$$\begin{aligned} & s^2 + 3.2\omega_0 s + \omega_0^2 \\ & s^3 + 1.75\omega_0 s^2 + 3.25\omega_0^2 s + \omega_0^3 \\ & s^4 + 2.41\omega_0 s^3 + 4.93\omega_0^2 s^2 + 5.14\omega_0^3 s + \omega_0^4 \\ & s^5 + 2.19\omega_0 s^4 + 6.50\omega_0^2 s^3 + 6.30\omega_0^3 s^2 + 5.24\omega_0^4 s + \omega_0^5 \\ & s^6 + 6.12\omega_0 s^5 + 13.42\omega_0^2 s^4 + 17.16\omega_0^3 s^3 + 14.14\omega_0^4 s^2 + 6.76\omega_0^5 s + \omega_0^6 \end{aligned}$$

By permission from *Transactions of the AIEE*, Volume 72, Part II, 1953, from article entitled "The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms", by Dunstan Graham and R. C. Lathrop.

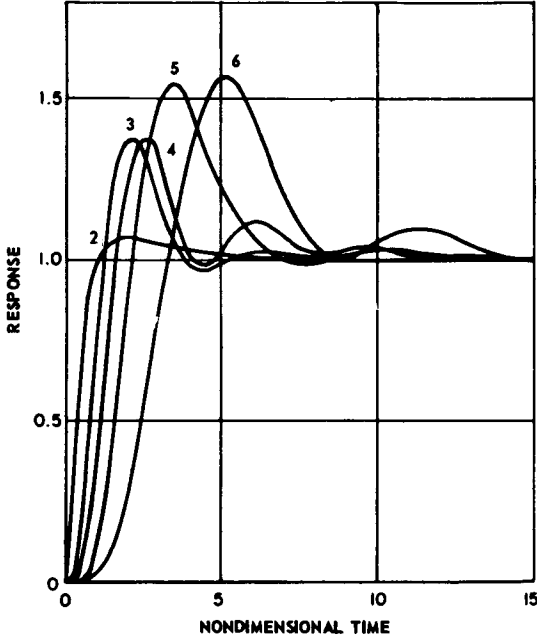


Fig. 8-4 Step-function responses of the optimum zero-velocity-error systems, second to sixth orders.

By permission from *Transactions of the AIEE*, Volume 72, Part II, 1953, from article entitled "The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms", by Dunstan Graham and R. C. Lathrop.

8-2.2 STATIONARY STOCHASTIC INPUTS (8,12,15)

The mean-square error (MSE) criterion is universally used as a performance index when the input is stochastic. The general configuration that applies to this minimization problem is shown in Fig. 8-5.

In this figure, $v(t)$ is the data component of the input, and $n(t)$ is the noise component of the input. The mean-square error is defined as

$$\overline{y_e^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} y_e^2(t) dt \quad (8-10)$$

On the assumption that the data and the noise are uncorrelated, application of the formulae given in Par. 3-8 to the configuration of Fig. 8-5 yields

$$\overline{y_e^2} = \int_{-\infty}^{+\infty} \Phi_{yy}(\omega) d\omega \quad (8-11)$$

where

$$\begin{aligned} \Phi_{yy}(s) &= [1 - W(s)] [1 - W(-s)] \\ \Phi_{vv}(s) + W(s) W(-s) \Phi_{nn}(s) \end{aligned} \quad (8-12)$$

$$W(s) = \frac{G_c(s) G_f(s)}{1 + G_c(s) G_f(s)} \quad (8-13)$$

$\Phi_{yy}(s)$ = power spectrum of system error y_e

$\Phi_{vv}(s)$ = power spectrum of data v

$\Phi_{nn}(s)$ = power spectrum of noise n

Since $\Phi_{yy}(\omega)$ is an even function, the evaluation of the integral in Eq. (8-11) can be carried out by means of the integral table (Table 8-1).

In all other respects, the design procedure for minimizing the mean-square error for stationary stochastic inputs with a fixed system configuration parallels the procedure for transient inputs outlined above.

Example. For the configuration of Fig. 8-5,

$$G_f(s) = \frac{1}{s(T_f s + 1)(T_m s + 1)}$$

$$G_c(s) = K_v$$

$$\Phi_{vn}(s) = 0$$

$$\Phi_{vv}(s) = \frac{\sigma_v^2 v / \pi}{-s^2(v^2 - s^2)}$$

$$\Phi_{nn}(s) = \frac{\gamma_n}{\pi} \text{ (white noise)}$$

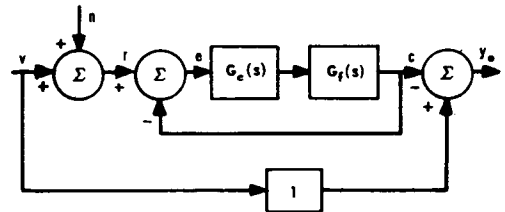


Fig. 8-5 Configuration for MSE minimization.

OPTIMIZATION METHODS

$$T_f = 0.01 \text{ sec}$$

$$T_m = 0.04 \text{ sec}$$

$$\sigma_v = 10 \text{ milliradian/sec}$$

$$v = 0.1 (\text{sec})^{-1}$$

$$\gamma_n = 0.4 (\text{milliradian})^2 - \text{sec}$$

If we assume that there is no noise initially ($\Phi_{nn} = 0$), then

$$1 - W(s) = \frac{s(T_f s + 1)(T_m s + 1)}{T_f T_m s^3 + (T_f + T_m)s^2 + s + K_v}$$

A normalized frequency is chosen such that, for $s = j\omega$, $T_f \omega = u$. By applying the normalization theorem of Par. 3-4 (Eq. 3-46) to the expression for $1 - W(s)$ above, we obtain

$$1 - W(ju) = \frac{(1 + ju)(1 + bju)}{ju b (ju)^3 + (1 + b)(ju)^2 + (ju) + K}$$

where

$$b = \frac{T_m}{T_f} = 4$$

$$K = T_f K_v$$

In addition, the power spectrum of data expressed in terms of u is found to be

$$\Phi_{vv}(u) = \left[\frac{\sigma_v^2 T_f^3}{\pi} \right] \left[\frac{a}{u^2(a^2 + u^2)} \right]$$

where

$$a = T_f v = 0.001$$

From Eq. (8-12), the power spectrum of system error expressed in terms of u is found to be

$$\Phi_{yy}(u) = \left[\frac{a \sigma_v^2 T_f^3}{\pi} \right] \left[\frac{c(u)}{D_4(ju) D_4(-ju)} \right]$$

where

$$C(u) = b^2 u^4 + (1 + b^2) u^2 + 1$$

$$D_4(ju) = b(ju)^4 + (1 + b + ab)(ju)^3 + (1 + a + ab)(ju)^2 + (a + K)(ju) + aK$$

$$D_4(-ju) = b(-ju)^4 + (1 + b + ab)(-ju)^3 + (1 + a + ab)(-ju)^2 + (a + K)(-ju) + aK$$

Then, using Eq. (8-11), we find the mean square error to be

$$\overline{y_e^2} = 2a\sigma_v^2 T_f^2 \frac{1}{2\pi} \left\{ \int_{-j\infty}^{+j\infty} \frac{c(u)}{D_4(ju) D_4(-ju)} du \right\}$$

Evaluating this expression from I_4 of the integral tables (Table 8-1) and substituting numerical values, we find that

$$\left[\frac{\sqrt{\overline{y_e^2}}}{\sigma_v T_f} \right]^2 N = \frac{0.016 K^2 - 3.99 K + 5.015}{-4K^3 + 5K^2 + 0.005 K}$$

To determine the minimum value of N it is convenient to make a plot of N versus K . This avoids a differentiation of N with respect to K which results in a fourth-degree algebraic equation whose roots must then be determined. It is evident that using the plot to determine the minimum is a simpler technique. The minimum from the plot is found to occur at $K = 1.1$ or $N = 0.90$. Consequently, the optimum system has

$$K_v = 90 \text{ sec}^{-1}$$

$$\left[\frac{\overline{y_e^2}}{\sigma_v^2} \right]^{1/2} = 0.095 \text{ milliradian}$$

If the noise is considered, the procedure is more involved but unchanged in principle. The results of the minimization of $\overline{y_e^2}$ with the noise added to the data as follows:

$$K_v = 7.8 \text{ sec}^{-1}$$

$$\left[\frac{\overline{y_e^2}}{\sigma_v^2} \right]^{1/2} = 2.21 \text{ milliradian}$$

8-3 OPTIMUM SYNTHESIS OF FREE-CONFIGURATION SYSTEMS WITH STATIONARY STOCHASTIC INPUTS ^(6,7,11,12,13,15)

The design problem for the optimum synthesis of free-configuration systems with stationary stochastic inputs is one of determining the closed-loop transfer function

$$\frac{C(s)}{R(s)} = W(s) \quad (8-14)$$

that minimizes the mean-square error when information is given concerning the data $v(t)$, noise $n(t)$, desired output $i(t)$, and fixed-element transfer function $G_f(s)$. No information concerning the form of the compensation $G_c(s)$ is needed. Figure 8-6 shows the configuration that describes the problem (G_d is the ideal element transfer function).

The solution to this problem is obtained by means of the calculus of variations and is in the form of an integral equation:

$$\begin{aligned} & \int_{-\infty}^{\infty} dt_3 g_f(t_3) \int_{-\infty}^{\infty} dt_4 g_f(t_4) \\ & \int_{-\infty}^{\infty} dt_2 \omega(t_2) \phi_{rr}(t_1 + t_3 - t_2 - t_4) \\ & - \int_{-\infty}^{\infty} dt_3 g_f(t_3) \phi_{ri}(t_1 + t_3) = 0 \text{ for } t_1 \geq 0 \end{aligned} \quad (8-15)$$

where

$g_f(t)$ = impulse response of fixed elements

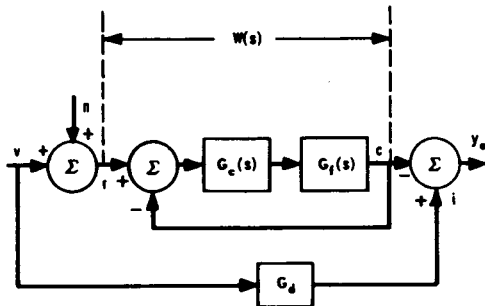


Fig. 8-6 Configuration for MSE minimization

$w(t)$ = impulse response of control system ($\mathcal{L}^{-1}[W(s)]$)

$\phi_{rr}(t)$ = autocorrelation function of $r(t)$

$\phi_{ri}(t)$ = crosscorrelation function between $r(t)$ and $i(t)$

If the fixed elements are minimum phase (no zeros in right-half s -plane) or are unspecified, Eq. (8-15) reduces to the Wiener-Hopf equation: ^(6,12,15)

$$\begin{aligned} & \int_{-\infty}^{\infty} dt_2 \omega(t_2) \phi_{rr}(t_1 - t_2) - \phi_{ri}(t_1) = \\ & \quad 0 \text{ for } t_1 \geq 0 \end{aligned} \quad (8-16)$$

If the autocorrelation functions and the impulse response of the fixed elements are Fourier transformable, Eqs. (8-15) and (8-16) can be solved in terms of transforms by a method called spectrum factorization. For Eq. (8-15), the optimum system function $W(s)$ is given by

$$W(s) = \left[\frac{\Gamma(s)}{\Delta^-(s)} \right]_+ \quad (8-17)$$

where

$$\Gamma(s) = 2\pi G_f(-s) \Phi_{ri}(s)$$

$$\Delta(s) = 2\pi G_f(s) G_f(-s) \Phi_{rr}(s)$$

$\Delta^+(s) \triangleq$ that factor of $\Delta(s)$ which contains all the poles and zeros of $\Delta(s)$ which lie in the left-hand s -plane

$$\Delta^-(s) = \frac{\Delta(s)}{\Delta^+(s)}$$

$\left[\frac{\Gamma(s)}{\Delta^-(s)} \right]_+ \triangleq$ that part of the *partial-fraction expansion* of $\Gamma(s)/\Delta^-(s)$ due to the poles of $\Gamma(s)/\Delta^-(s)$ which lie in the left-half s -plane

$$\left[\frac{\Gamma(s)}{\Delta^-(s)} \right]_- = \frac{\Gamma(s)}{\Delta^-(s)} - \left[\frac{\Gamma(s)}{\Delta^-(s)} \right]_+$$

Transformation and factorization of Eq. (8-16) yields

$$W(s) = \frac{\left[\frac{\Phi_{rr}(s)}{\Phi_{rr}^-(s)} \right]}{\Phi_{rr}^+(s)} \quad (8-18)$$

where the notation is the same as that defined below Eq. (8-17).

Example. $G_v(s)$ is minimum phase, and

$$\Phi_{vv}(s) = \frac{\sigma_v^2 v}{\pi(-s^2)(v^2 - s^2)}$$

$$\Phi_{nn}(s) = \frac{\gamma_n}{\pi}$$

$$\Phi_{vn}(s) = 0$$

$$\sigma_v = 10 \text{ milliradian/sec}$$

$$v = 0.1 \text{ sec}^{-1}$$

$$\gamma_n = 0.4 \text{ (milliradian)}^2 - \text{sec}$$

$$i(t) = v(t)$$

Normalize the frequency scale by letting

$$\lambda = \frac{s}{v}$$

Then,

$$\Phi_{vv}(\lambda) = \frac{\sigma_v^2}{\pi v^2} \frac{1}{(-\lambda^2)(1 - \lambda^2)}$$

$$\Phi_{nn}(\lambda) = \frac{v \gamma_n}{\pi}$$

Let

$$\frac{\sigma_v^2}{v^2} = a \quad \text{and} \quad v \gamma_n = b$$

Since the data and noise are uncorrelated

$$(\Phi_{vn} = 0),$$

$$\Phi_{rr}(\lambda) = \Phi_{vv}(\lambda) + \Phi_{nn}(\lambda) = \frac{b}{\pi} \left[\frac{c^2 - \lambda^2 + \lambda^4}{(-\lambda^2)(1 - \lambda^2)} \right]$$

where

$$c^2 = \frac{a}{b} = \frac{\sigma_v^2}{v^3 \gamma_n} = 25 \times 10^4$$

Since the desired output $i(t)$ is the data $v(t)$,

$$\Phi_{ri}(\lambda) = \Phi_{vv}(\lambda)$$

Equation (8-18) applies to this problem. To find $\Phi_{rr}(\lambda)$ and $\Phi_{rr}^+(\lambda)$ it is necessary to distinguish between poles and zeros in the two half planes. Since $\Phi_{rr}(\lambda)$ has a double pole on the imaginary axis at the origin, the following artifice is used. We let

$$-\lambda^2 = \lim_{\epsilon \rightarrow 0} (\epsilon - \lambda)(\epsilon + \lambda)$$

The problem can then be worked with $(\epsilon - \lambda)(\epsilon + \lambda)$ replacing $-\lambda^2$. After carrying out the pertinent algebraic manipulations, we let $\epsilon \rightarrow 0$. This is equivalent to factoring $-\lambda^2$ into $(-\lambda)(+\lambda)$ and then associating $(-\lambda)$ with the right-half plane (RHP) and $(+\lambda)$ with the left-half plane (LHP). Therefore, the factorization of $\Phi_{rr}(\lambda)$ becomes

$$\Phi_{rr}(\lambda) = \left[\frac{(m + jn + \lambda)(m - jn + \lambda)}{(+\lambda)(1 + \lambda)} \right] \left[\frac{b}{\pi} \frac{(m + jn - \lambda)(m - jn - \lambda)}{(-\lambda)(1 - \lambda)} \right]$$

where

$$m = 0.5 \sqrt{2c + 1} = 15.82$$

$$n = 0.5 \sqrt{2c - 1} = 15.80$$

The factor of $\Phi_{rr}(\lambda)$ having all its poles and zeros in the right-half plane is

$$\Phi_{rr}^-(\lambda) = \frac{b}{\pi} \frac{(m + jn - \lambda)(m - jn - \lambda)}{(-\lambda)(1 - \lambda)}$$

Therefore,

$$\frac{\Phi_{ri}(s)}{\Phi_{rr}^-(s)} = \frac{c^2}{(+\lambda)(1 + \lambda)(m + jn - \lambda)(m - jn - \lambda)}$$

This function has left-half-plane poles at $\lambda = 0$ and $\lambda = -1$. Expanding in terms of partial fractions and retaining only those terms in the expansion due to LHP poles, we obtain

THEORY

$$\left[\frac{\Phi_{rr}(\lambda)}{\Phi_{rr}^-(\lambda)} \right]_+ = c \left\{ \frac{1}{\lambda} - \frac{\frac{c}{1+c+\sqrt{2c+1}}}{1+\lambda} \right\}$$

or

$$\left[\frac{\Phi_{rr}(\lambda)}{\Phi_{rr}^-(\lambda)} \right]_+ = (\sqrt{2c+1} - 1) \frac{\left(\frac{1+\sqrt{2c+1}}{2} + \lambda \right)}{+\lambda(1+\lambda)}$$

Since,

$$\Phi_{rr}^+(\lambda) = \frac{(m+jn+\lambda)(m-jn+\lambda)}{(+\lambda)(1+\lambda)}$$

$$W(\lambda) = (\sqrt{2c+1} - 1) \frac{\left(\frac{1+\sqrt{2c+1}}{2} + \lambda \right)}{(\lambda^2 + \sqrt{2c+1}\lambda + c)}$$

Numerically, since $\lambda = 10s$ and $c = 500$, the optimum response is

$$W(s) = \frac{1 + 0.613s}{\left(\frac{s}{\omega_n} \right)^2 + 2\zeta \left(\frac{s}{\omega_n} \right) + 1}$$

where

$$\zeta = \sqrt{\frac{1}{2} + \frac{1}{4c}} = 0.707$$

$$\omega_n = v\sqrt{c} = 2.24$$

The open-loop transfer function corresponding to the optimum response is

$$\frac{C(\lambda)}{E(\lambda)} = G_c(\lambda) G_f(\lambda) = c \frac{\left(1 + \frac{2}{1+\sqrt{2c+1}}\lambda \right)}{\lambda(1+\lambda)}$$

or, numerically,

$$\frac{C(s)}{E(s)} = 50 \frac{(0.613s + 1)}{s(10s + 1)}$$

The mean-square error $\overline{y_e^2}$ can then be evaluated by using Eqs. (8-11) and (8-12) together with Table 8-1. The rms error due to noise alone is found to be

$$\left[\overline{y_e^2} \right]_n^{1/2} = 1.35 \text{ milliradian}$$

The rms error due to the signal acting alone is

$$\left[\overline{y_e^2} \right]_v^{1/2} = 0.796 \text{ milliradian}$$

Consequently, the total rms error is

$$\left[\overline{y_e^2} \right]^{1/2} = \sqrt{\left[\overline{y_e^2} \right]_n + \left[\overline{y_e^2} \right]_v}$$

or

$$\left[\overline{y_e^2} \right]^{1/2} = 1.57 \text{ milliradian}$$

8-4 LIMITATIONS AND APPLICATION PROBLEMS

Several difficulties confront the designer who carries out the optimization procedure in any practical problem. He finds that (1) the labor involved becomes excessive, and (2) the resulting optimum compensation is both difficult to realize and unrealistic. The latter difficulty arises because cancellation of the fixed-component transfer function is required, resulting in component saturation and poor utilization of hardware. One factor acting in favor of the designer using the ISE and MSE criteria is that the minima resulting from the use of these procedures are broad. Thus, a fairly wide deviation of parameters and configurations can occur without appreciably altering the performance index. Hence, the labor involved in designing by optimization techniques can be reduced greatly by judicious engineering approximations. In addition, more freedom is available to the designer when the minima that arise in an optimization problem are broad since he can then satisfy additional performance specifications such as rise time, peak overshoot, etc. The same cannot be said for the ITAE criterion since it is a selective criterion producing narrow minima that leave little freedom to the designer. Therefore, the ITAE criterion is not to be recommended if parameter variation is to be expected and other performance specifications are to be met.

Techniques for overcoming the second limitation of optimization procedures have been proposed by Newton.^(7,10) He recommends that constraints be placed on the signals that are not to saturate. That is, the optimization is to be carried out by requiring that the performance index be minimized while the signals that may saturate are kept below a specified upper limit. Actually, however, a measure of the peak-signal values is used to facilitate

analysis. In the case of transient inputs, the integral-square signal values are to be kept below assigned limits. In the case of stochastic signals, the rms signal values are to be constrained. It is also possible to combine the two types of signals by requiring, for example, that the rms error for a stochastic input be minimized subject to a constraint on the integral-square value of a specified signal for a transient input.

Newton⁽¹⁰⁾ also proposes that constrained optimization be carried out by minimizing bandwidth since a minimum bandwidth system is highly desirable in any case. Thus, bandwidth is minimized subject to a constraint on the allowable error index.

By employing constrained optimization using performance indices that exhibit broad minima, the designer can approach a problem with a greater degree of certainty of finding out whether his specifications are compatible and, if compatible, whether they can be met in practice.

The optimization procedures discussed here have been limited to transient inputs and stationary stochastic inputs. If the input is nonstationary, then the optimum system will become nonstationary or time-variable. If the form of the time variation in the input statistics is known, it is possible to design a system which exhibits variable bandwidth. Unfortunately, there is as yet no general method for finding an explicit analytical solution to this problem. If the time variation in the input statistics is slow compared with the response time of the system, then the nonstationary problem can be broken down into a series of stationary segments. If the input statistics vary at a rate that is of the order of the response time of the system, then one cannot ignore the nonstationary nature of the problem.

BIBLIOGRAPHY

- 1 P. T. Nims, "Some Design Criteria for Automatic Controls", *Trans. AIEE*, Vol. 70, Part I, pp. #606-611, 1951.
- 2 F. C. Fickeison and T. M. Stout, "Analogue Methods for Optimum Servomechanism Design", *Trans. AIEE*, Vol. 71, Part II, pp. #244-250, 1952.
- 3 D. Graham and R. C. Lathrop, "The Synthesis of 'Optimum' Transient Response: Criteria and Standard Forms", *Trans. AIEE*, Vol. 72, Part II, pp. #273-288, 1953.
- 4 R. C. Lathrop and D. Graham, "The Transient Performance of Servomechanisms with Derivative and Integral Control", *Trans. AIEE*, Vol. 73, Part II, pp. #10-17, 1954.
- 5 D. Graham and R. C. Lathrop, "The Influence of Time Scale and Gain on Criteria for Servomechanism Performance", *Trans. AIEE*, Vol. 73, Part II, pp. #153-158, 1954.
- 6 N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley & Sons, Inc., New York, N. Y., 1949.
- 7 G. C. Newton, Jr., "Compensation of Feedback Control Systems Subject to Saturation", *J. Franklin Inst.*, Vol. 254, pp. #281-296, 391-413, October, November, 1952.
- 8 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms*, MIT Radiation Laboratory Series, Vol. 25, pp. #262-370, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.
- 9 J. H. Westcott, "The Introduction of Constraints into Feedback System Designs", *Trans. IRE*, CT-1, pp. #39-49, September, 1954.
- 10 G. C. Newton, Jr., "Design of Control Systems for Minimum Bandwidth", *Trans. AIEE*, Vol. 74, Part II, pp. #161-168, 1955.
- 11 H. S. Tsien, *Engineering Cybernetics*, pp. #111-135, 231-252, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.
- 12 J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. #410-499, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.
- 13 H. W. Bode and C. E. Shannon, "A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory", *Proc. IRE*, Vol. 38, pp. #417-425, April, 1950.
- 14 L. A. Zadeh and J. R. Ragazzini, "An Extension of Wiener's Theory of Prediction", *J. Appl. Phys.*, Vol. 21, pp. #645-655, July, 1950.
- 15 J. H. Laning, Jr. and R. H. Battin, *Random Processes in Automatic Control*, McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

CHAPTER 9

SAMPLED-DATA SYSTEMS***9-1 GENERAL THEORY**

Linville states, "A sampled-data control system is one wherein the signal supplied to one or more parts of the system is not given continuously in time, but is supplied at discrete values of the time variable, t . In such a system, the part of the system being fed intermittently might, for example, have an input signal applied to it at $t = 0, T, 2T, 3T, \dots$ (where T is the length of time between samplings) with no data at all supplied in the intervals separating these sampling instants. A control system makes use of sampled data when it is impossible to supply continuous data to all its parts."[†]

[†]Quoted by permission from *Transactions of the AIEE*, Volume 70, Part II, 1951, from article entitled "Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation," by W. K. Linville.

If the sampling frequency is high compared to the signal frequency and the critical frequencies of the system, then the fact that the data are sampled has little bearing on system behavior. Otherwise, the effect of sampling may become quite pronounced.

Figure 9-1 shows the elements of a typical sampled-data system. The input $r(t)$ may be composed of sampled or continuous data. The sampling device periodically samples the actuating signal $e(t)$ under control of the carrier signal supplied to it. The holding circuit is used to smooth the sampled output from the sampling device, and the smoothed output of the holding circuit then drives the output member. It is evident that the components and signals in the system are combinations of discrete and continuous elements. Because part

*By L. A. Gould

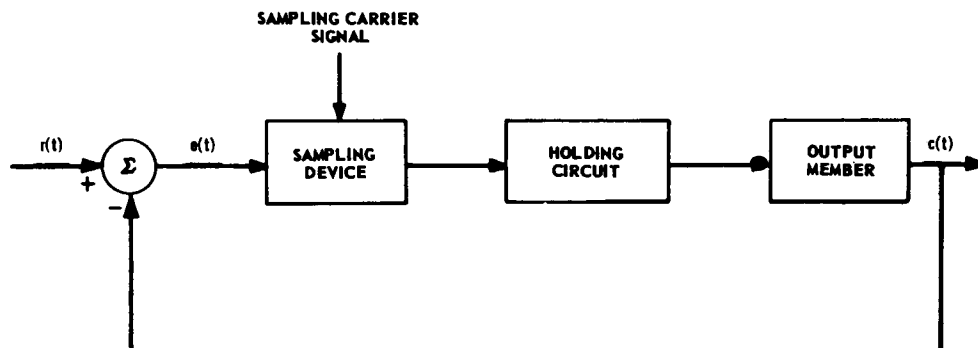


Fig. 9-1 Sampled-data system.

THEORY

of the system operates on sampled data and part on continuous data, the analysis of system behavior is not easily carried out by conventional methods. For that part of the system operating on continuous data, conventional methods of analysis are best. For that part operating on sampled data, the use of sequences and linear difference equations is best. However, methods have been developed which treat sampled-data systems from a unified viewpoint and will be presented in this chapter.

A sampled-data system like that in Fig. 9-1 can be represented by the mathematical model shown in Fig. 9-2. The impulse modulator is an ideal device that multiplies the actuating signal $e(t)$ by the carrier signal $\Delta(t)$. The function $\Delta(t)$ represents a periodic train of unit impulses occurring at a frequency $\Omega = 2\pi/T$ radians per second (Fig. 9-3)—where Ω is called the sampling frequency. The equivalent linear filter is so chosen that the combined operation of the impulse modulator and equivalent linear filter on the actuating signal in the model produces the same input to the output member as the combined action of the sampling device and holding circuit in the original system. For example, a system in which the actuating signal is sampled every T seconds by a device which holds a particular

sample value at that value until the next sampling time is called a sampler-clamper. Its effect is shown in Fig. 9-4. This type of behavior can be exactly represented by the combination of an impulse modulator and a filter whose transfer function $G_e(s)$ is

$$G_e(s) = \frac{1 - e^{-st}}{s} \quad (9-1)$$

In the mathematical model, the output of the impulse modulator will be a train of impulses, the magnitude of each being the value of the actuating signal at the corresponding sampling time. Although such a signal does not exist in the physical system, it is useful to isolate the action of the impulse modulator, and combine the equivalent linear filter with the output member for the purpose of analysis. The impulse modulator is thought of as a synchronous switch, controlled by a carrier, which periodically closes the connection between the actuating signal and the input to the equivalent linear filter (Fig. 9-5). The signal $e(t)$ entering the switch is continuous in this picture, but the signal $e^*(t)$ leaving the switch is discrete (sampled). It is also possible for the signal $e(t)$ to be discrete as well, in which case the action of the switch has no effect if it is synchronized with the discrete intervals associated with the input. We will adopt the convention that an impulse modulator (or synchronous switch) operates on all

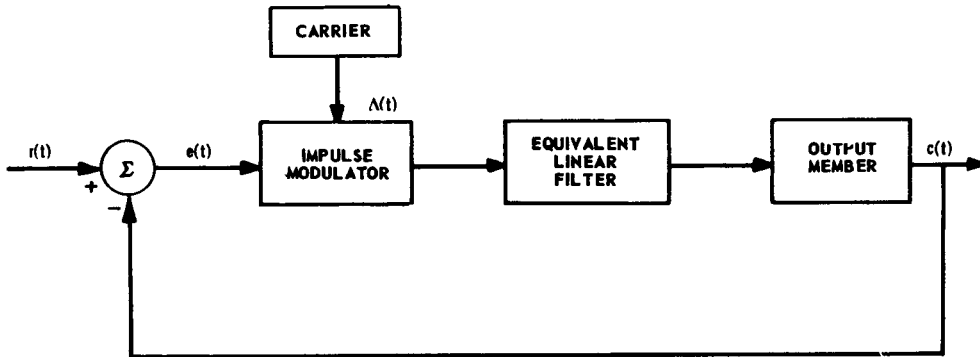


Fig. 9-2 Model of sampled-data system.

SAMPLED-DATA SYSTEMS

signals entering it (whether continuous, discrete, or a combination of both) to produce a discrete output. Signals that have been sampled, and are therefore discrete, will be represented by starred functions. Thus, the operation of an impulse modulator is as represented in Fig. 9-6.

The carrier signal $\Delta(t)$ in a sampled-data system is represented by

$$\Delta(t) = \sum_{n=-\infty}^{+\infty} \delta_o(t - nT) \quad (9-2)$$

The Laplace transform of this function is

$$\Delta(s) = \sum_{n=0}^{+\infty} e^{-nTs} \quad (9-3)$$

or

$$\Delta(s) = \frac{1}{1 - e^{-sT}} \quad (9-4)$$

If a signal $r(t)$ is sampled by an impulse modulator, the sampled output $r^*(t)$ is

$$r^*(t) = \Delta(t) r(t) \quad (9-5)$$

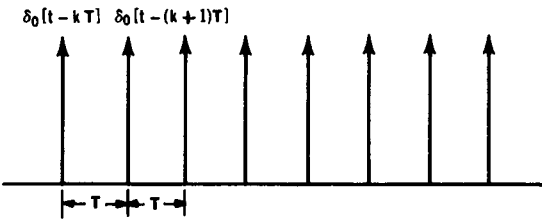


Fig. 9-3 Train of unit impulses which represents the carrier $\Delta(t)$.

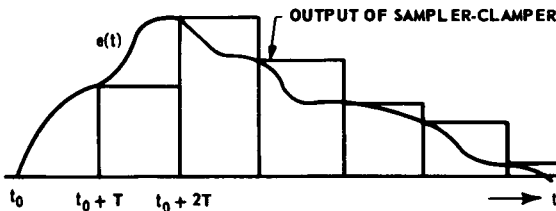


Fig. 9-4 Action of sampled-clamper.

or

$$r^*(t) = \sum_{n=-\infty}^{+\infty} r(nT) \delta_o(t - nT) \quad (9-6)$$

The Laplace transform $R^*(s)$ of a sampled signal is

$$R^*(s) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} R(s + jn\Omega) \quad (9-7)$$

where

$$\Omega = \frac{2\pi}{T} \quad (9-8)$$

and

$$R(s) = \mathcal{L}[r(t)] \quad (9-9)$$

Another form of the transform of a sampled signal is

$$R^*(s) = \sum_{n=0}^{\infty} r(nT) e^{-nTs} \quad (9-10)$$

Note that a starred transform like $R^*(s)$ represents the transform of a starred (sampled) time function. Also from Eq. (9-7), starred transforms are periodic functions of frequency, the period being $j\Omega$. That is,

$$R^*(s) = R^*(s + jn\Omega) \quad (9-11)$$

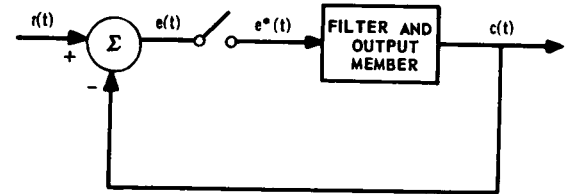


Fig. 9-5 Simplified picture of sampled-data system.

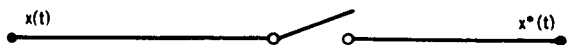


Fig. 9-6 Operation of sampling switch.

9-2 THE z TRANSFORM AND THE w TRANSFORM

9-2.1 THE z TRANSFORM

Whenever a time function is transformable, it can be shown that the transform of the function when sampled is a rational function of e^{-Ts} ; i.e.,

$$R^*(s) = F(e^{-sT}) \quad (9-12)$$

where $R^*(s)$ is the transform of the sampled function $r^*(t)$. If we let

$$z = e^{-sT} \quad (9-13)$$

then transforms of sampled-time functions are functions of a new complex variable z . The z transform of a time function $r(t)$ is then defined as the Laplace transform $R^*(s)$ of $r^*(t)$, where $r^*(t)$ is the function produced by impulse modulating $r(t)$, and the z transform is obtained by replacing e^{-sT} by z in $R^*(s)$. It is conventional to let $R^*(z)$ represent the z transform of $r(t)$ although, rigorously, one should use $R^*(-1/T \log_e z)$.

If the Laplace transform $R(s)$ of a function $r(t)$ is known, the z transform $R^*(z)$ can be found by expanding $R(s)$ in a partial-fraction expansion and using the formulae given below. If

$$R(s) = \sum_{k=1}^m \frac{K_k}{s + a_k} \quad (9-14)$$

then

$$R^*(z) = \sum_{k=1}^m \frac{K_k}{1 - ze^{-a_k T}} \quad (9-15)$$

A short table of z transforms and their equivalent Laplace transforms is given in Table 9-1. For more extensive tables see references 5, 25, and 26. Unfortunately, several authors have adopted the relation $z = e^{+sT}$. This notation arose from the mathematics of difference equations, but it is awkward and physically deceiving in the present connection, since e^{+sT} corresponds to ideal prediction. Therefore, when using the literature, care must be taken to verify which particular notation is being used. In reference 5, for example, all the

expressions for z transforms should have z replaced by z^{-1} to make them correspond to the notation adopted in this chapter.

The introduction of the z transform enables one to treat sampled-data systems by all the techniques available for continuous-data systems since it is evident that the process of sampling a time function can be represented by a

TABLE 9-1
LAPLACE AND z TRANSFORM PAIRS

| | Laplace Transform: $F(s)$ | z Transform: $F^*(z)$ |
|-----|--------------------------------------|--|
| 1. | 1 | 1 |
| 2. | e^{-sT} | z^{-1} |
| 3. | $\frac{1}{s}$ | $\frac{1}{1 - z}$ |
| 4. | $\frac{1}{s^2}$ | $\frac{Tz}{(1 - z)^2}$ |
| 5. | $\frac{1}{s + a}$ | $\frac{1}{1 - ze^{-aT}}$ |
| 6. | $\frac{a}{s(s + a)}$ | $\frac{z(1 - e^{-aT})}{(1 - z)(1 - ze^{-aT})}$ |
| 7. | $\frac{a}{s^2 + a^2}$ | $\frac{z \sin aT}{1 - 2z \cos aT + z^2}$ |
| 8. | $F(s + a)$ | $F^*(ze^{-aT})$ |
| 9. | $e^{-sT}F(s)$ | $zF^*(z)$ |
| 10. | $e^{as}F(s)$ | $z^{-(a/T)}F^*(z)$ |
| 11. | $\frac{1}{s + \frac{1}{T} \log_e a}$ | $\frac{1}{1 - az}$ |
| 12. | $\frac{s}{s^2 + a^2}$ | $\frac{1 - z \cos aT}{1 - 2z \cos aT + z^2}$ |
| 13. | $\frac{1}{(s + a)^2}$ | $\frac{Tze^{-aT}}{(1 - ze^{-aT})^2}$ |

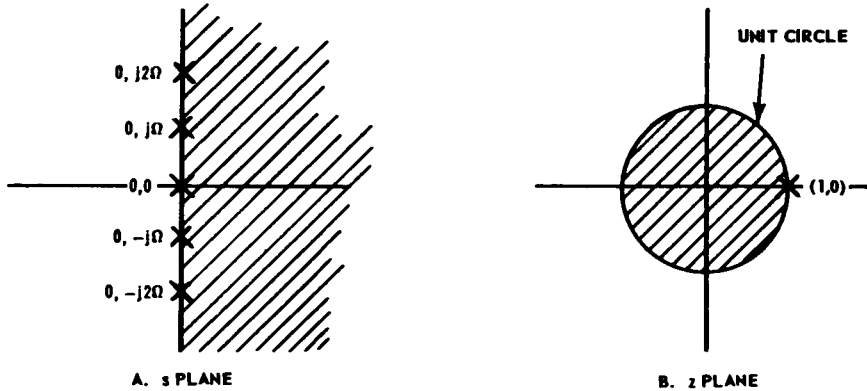


Fig. 9-7 Relations between s plane and z plane.

variable transformation from the s domain to the z domain. The z transform is a conformal transformation that maps the right half of the s plane into the interior of the unit circle of the z plane and the left half of the s plane into the exterior of the unit circle of the z plane. The imaginary axis of the s plane is mapped into the unit circle of the z plane, the s -plane origin $(0 + j0)$ mapping into the point $(+1 + j0)$ in the z plane. These relations are shown in Fig. 9-7. Because a z transform is periodic (period $= j\Omega$), the points $(0 + jk\Omega)$ ($k = 1, 2, 3, \dots$) in the s plane also map into the point $(+1 + j0)$ in the z plane as shown in Fig. 9-7. Similar relationships are easily visualized and are treated later in this chapter. There is one point which must be emphasized, however. Since the z transform of a time function represents the Laplace transform of the corresponding sampled-time function, information about behavior between sampling instants is lost and cannot be recovered from inspection of z transforms. However, Barker⁽²⁴⁾ has developed a method for determining behavior between sampling instants. This method is described in Par. 9-5. It should also be noted that the z transform is related to the Mellin transform which is used to develop the theory of transforms and to study problems in elasticity.⁽²⁸⁾

9-2.2 THE w TRANSFORM

Because of the difficulty involved in relating some of the properties of sampled-data systems to the frequency-domain concepts that are most convenient to apply in the study of continuous systems, Johnson et al⁽¹⁴⁾ have suggested a very useful transformation of variables that aids greatly in design. If

$$z = \frac{1 - w}{1 + w} \quad (9-16)$$

is a bilateral transformation from the z plane to the w plane, then

$$R^*(w) = R^*(z) \quad \left| \quad z = \frac{1 - w}{1 + w} \right| \quad (9-17)$$

is defined as the w transform of $r(t)$. The advantage of introducing the w transform becomes evident when an attempt is made to evaluate $R^*(s)$ for $s = j\omega$. Such an evaluation requires an infinite "vector" sum, theoretically [Eq. (9-7)], or else evaluation of R^* through the use of $e^{-j\omega T} = \cos \omega T + j \sin \omega T$ and so it is fairly difficult to obtain in practice. The use of the w transform, on the other hand, simplifies the determination of the frequency response of sampled-data systems. The w transform maps the unit circle in the z plane

THEORY

into the imaginary axis in the w plane and restores some of the analytical advantages that were lost through the sampling process. If

$$w = u + jv \quad (9-18)$$

then

$$v = \tan \left(\frac{\omega T}{2} \right) \quad (9-19)$$

gives the relation between the real frequency ω and the pseudo-frequency v corresponding

to the imaginary part of w . The real frequency ω , from Eq. (9-19), is

$$\omega = \frac{2}{T} \tan^{-1} v \quad (9-20)$$

The primary advantage of the w transform is that the transforms of sampled signals can be represented by a rational function of a frequency variable w that is simply related to the frequency ω . In the following sections this property is brought out clearly.

9-3 OPERATIONAL METHODS

9-3.1 GENERAL

The operational definition of impulse modulation given in Par. 9-1 simplifies the study of sampled-data systems.

9-3.2 BASIC RELATIONS OF SAMPLED FUNCTIONS

The following basic relations are easy to verify. From Fig. 9-8, it is evident that:

$$C(s) = G(s)R^*(s) \quad (9-21)$$

$$C^*(s) = G^*(s)R^*(s) \quad (9-22)$$

$$[R^*(s)]^* = R^*(s) \quad (9-23)$$

From Fig. 9-9, it is evident that:

$$C(s) = G(s)R(s) \quad (9-24)$$

$$C^*(s) = [G(s)R(s)]^* \quad (9-25)$$

$$C^*(s) \neq G^*(s)R^*(s) !! \quad (9-26)$$

From Fig. 9-10, it is evident that:

$$\frac{E^*(s)}{R^*(s)} = \frac{1}{1 + G^*(s)} \quad (9-27)$$

$$\frac{C(s)}{R^*(s)} = \frac{G(s)}{1 + G^*(s)} \quad (9-28)$$

$$\frac{C^*(s)}{R^*(s)} = \frac{G^*(s)}{1 + G^*(s)} \quad (9-29)$$

The foregoing relations indicate that all the techniques of block-diagram algebra can be used to manipulate sampled-data-system configurations except for the added restrictions

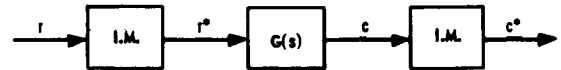


Fig. 9-8 Sampling a smoothed sampled signal.

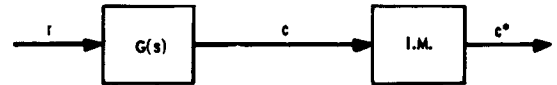


Fig. 9-9 Sampling a filtered continuous signal.

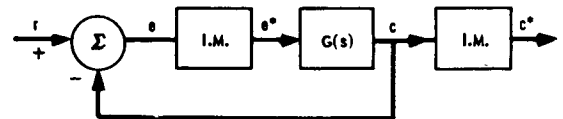


Fig. 9-10 A sampled-data feedback system.

that an impulse modulator (a) “stars” all signals entering it, and (b) its position in a diagram cannot be interchanged with a continuous transfer function. Equation (9-26) is included to emphasize the fact that the starred product of two Laplace transforms is *not* equal to the product of the corresponding starred transforms.

The equations relating to Fig. 9-10 [Eqs. (9-27) through (9-29)] introduce some of the properties of sampled-data feedback systems. In particular, the stability of a sampled-data system is related directly to the zeros of the following expression:

$$1 + G^*(s) = 0 \quad (9-30)$$

If any of the roots of this equation lie in the right half of the s plane, the system is unstable. Nyquist’s criterion (Par. 4-3) can be applied directly to determine the stability of sampled-data systems, except for one modification. Since $G^*(s)$ is a periodic function, it has an infinite number of poles and zeros, but the pole-zero configuration is repeated for every multiple of $j\Omega$. Similarly, the $G^*(s)$ locus in the s plane is repeated every time s changes by $j\Omega$. Because the $G^*(s)$ plot is symmetrical about the real axis in the s plane, $G^*(s)$ need only be plotted for $0 < j\omega < j\Omega/2$ when $s = j\omega$. In practice, the s -plane contour

and the corresponding $G^*(s)$ locus are as shown in Fig. 9-11 when the Nyquist criterion is applied. In terms of the z plane, Eq. (9-30) becomes

$$1 + G^*(z) = 0 \quad (9-31)$$

The stability of the system is determined by plotting $G^*(z)$ as z traverses the unit circle. If there are any roots of Eq. (9-31) that lie *inside* the unit circle, the system is unstable. The difficulty encountered in plotting $G^*(s)$ or $G^*(z)$ from the required variation of s or z is removed when the w transform is introduced. $G^*(w)$ is easily handled since it is expressible as a rational function of a frequency variable.

Example. Consider a simple servomechanism with block diagram shown in Fig. 9-10. The physical device includes a sampler-clamper (See Par. 9-1), a servomotor having a one-second time constant, and an ideal amplifier. The transfer function of the continuous portion [including the filter as in Eq. (9-1)] is given by Eq. (9-32).

$$G(s) = K \left(\frac{1 - e^{-s}}{s} \right) \frac{1}{s(s+1)}, \quad T = 1 \text{ sec.} \quad (9-32)$$

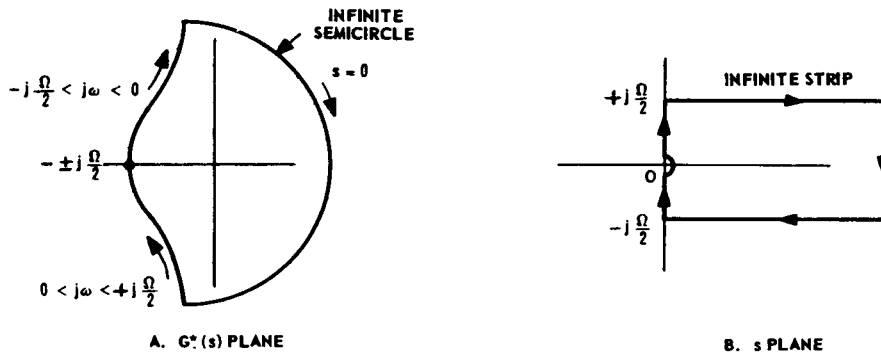


Fig. 9-11 Relations between s and $G^*(s)$ for application of Nyquist criterion.

THEORY

From Eq. (9-7), we obtain

$$\frac{1}{K} G^*(s) = \frac{1}{(1 - e^{-s}) \sum_{n=-\infty}^{+\infty} \frac{1}{(s + jn\Omega)^2 (1 + s + jn\Omega)}} \quad (9-33)$$

where $\Omega = 2\pi/T = 2\pi$.

This function is difficult to plot for $s = j\omega$. Taking the z transform of Eq. (9-32), we get

$$\frac{1}{K} G^*(z) = \frac{z(0.264z + 0.368)}{(0.368z - 1)(z - 1)} \quad (9-34)$$

This function is difficult to plot for $z = e^{-j\omega}$. Letting

$$z = \frac{1 - w}{1 + w} \quad (9-35)$$

the w transform of Eq. (9-32) is found to be

$$\frac{1}{K} G^*(w) = \frac{(1 - w)(0.632 + 0.104w)}{(0.632 + 1.368w)(2w)} \quad (9-36)$$

when $w = jv$, the function in Eq. (9-36) is easily handled by conventional techniques since the relation between the w plane and the $G^*(w)$ plane is the same kind of relation as that which exists between the s plane and the $G(s)$ plane. If the real frequency ω is to be considered, then there is an added difficulty in that Eq. (9-20) must be used to calibrate the frequency locus. The asymptotic and gain-phase plane techniques can be used with a change only in the relation between v and ω .

9-3.3 ADDITIONAL PROPERTIES OF SAMPLED FUNCTIONS

(a) $G^*(e^{-sT}) =$

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(p) \left[\frac{1}{1 - e^{-T(p-s)}} \right] dp \quad (9-37)$$

(b) If $G(s) = \frac{N(s)}{D(s)}$, and $D(s)$ has only simple poles,

$$G^*(e^{-sT}) = \sum_{n=1}^l \frac{A(s_n)}{B'(s_n)} \frac{1}{1 - e^{-T(s-s_n)}} \quad (9-38)$$

where s_n is the n th pole of $G(s)$.

(c) If the z transform of a time function $r(t)$ is given, the value of the time function at the sampling instants can be found from

$$r^*t = r(nT) = \frac{1}{2\pi j} \oint R^*(z) z^{t-n} dz \quad (9-39)$$

where the contour integration in the z plane is along a path that encloses all the singular points of $R^*(z)z^{t-n}$.

(d) Initial-Value Theorem:

$$\lim_{t \rightarrow 0} r(t) = \lim_{z \rightarrow 0} (1 - z) R^*(z) \quad (9-40)$$

(e) Final-Value Theorem: If $R^*(z)$ has all its poles outside the unit circle of the z plane,

$$\lim_{t \rightarrow \infty} r(t) = \lim_{z \rightarrow 1} (1 - z) R^*(z) \quad (9-41)$$

(f) For $s = \pm j\frac{k\Omega}{2}$ ($k = 0, 1, 2, \dots$), $G^*(z)$ is always real $\left(\Omega = \frac{2\pi}{T} \right)$.

(g) The degree of the denominator of $G^*(z)$ in z always equals the degree of the denominator of $G(s)$ in s if $G(s)$ is rational.

(h) The poles of $G^*(z)$ in the strip $-j\frac{\Omega}{2} < \text{Im}(s) < +j\frac{\Omega}{2}$ in the s plane are the poles of $G(s)$ in the s plane.

(i) Changing the values of the poles of $G(s)$ changes the coefficients $A(s_n)/B'(s_n)$ as well as the terms $1/(1 - ze^{+Ts_n})$ in the partial-fraction expansion of $G^*(z)$ [Eq. (9-38)].

(j) Insertion of zeros in $G(s)$ changes only the coefficients $A(s_n)/B'(s_n)$.

(k) The number of poles of $G^*(z)$ at $z = 1$ is equal to the number of poles of $R(s)$ at $s = 0$.

(l) In terms of the w transform, the initial value theorem is

$$\lim_{t \rightarrow 0} r(t) = \lim_{w \rightarrow 1} \left(\frac{2w}{1 + w} \right) R^*(w) \quad (9-42)$$

and the final-value theorem is

$$\lim_{t \rightarrow \infty} r(t) = \lim_{w \rightarrow 0} \left(\frac{2w}{1 + w} \right) R^*(w) \quad (9-43)$$

9-4 DESIGN TECHNIQUES

The problem of designing a sampled-data system is complicated by the fact that the system can contain both discrete and continuous elements. In addition, direct application of z -transform theory merely gives the response at the sampling instants, but the behavior during the sampling instants cannot be determined by simple methods.

The insertion of a sampling device in an otherwise continuous system to produce a sampled-data system introduces the following limitations:

(a) A greater tendency towards instability results.

(b) Ripple components arise in the output at the sampling frequency and its harmonics.

(c) The usable bandwidth of the system is reduced to a fraction of the sampling frequency Ω , the theoretical upper limit being $\Omega/2$.

To determine the gain necessary to stabilize the system for a specified M_p in the closed-loop frequency response, the introduction of the w transform greatly facilitates plotting the frequency locus as indicated in Par. 9-3. Conventional continuous-system techniques can be used.

The root-locus procedure can be used in a conventional manner in the z plane to investigate the closed-loop pole-zero configuration. This procedure differs from that used for s -plane loci of continuous systems in that: (a) instability implies closed-loop poles inside the unit circle of the z plane (as contrasted to right-half-plane poles in continuous-system design), and (b) the dominant pole (or pole pair) is the pole nearest the point (1,0) in the z plane (as contrasted to poles nearest the origin in the s plane). Otherwise, conventional procedures can be used to investigate stability, relative stability, and the effect of compensation.

Compensation of sampled-data systems with continuous networks (conventional lead and lag networks) is a difficult design problem and is best treated by trial-and-error analysis. In many important applications discrete networks can be used for compensation; for example, the use of digital computers in fire-control systems provides the designer with an opportunity to use digital (discrete) filters in the compensation of the control system. Figure 9-12 shows the difference between continuous and discrete com-

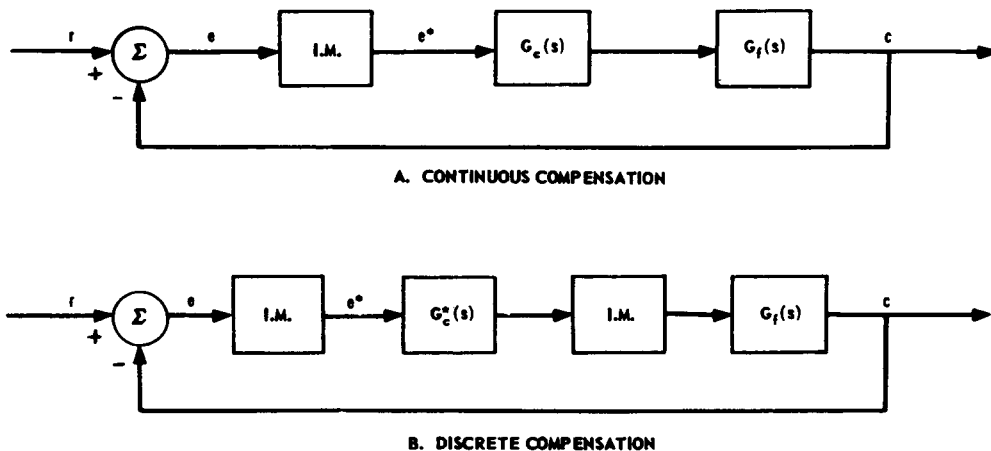


Fig. 9-12 Comparison between discrete and continuous compensation.

pensation. In Fig. 9-12A, $G_c(s)$ is the transfer function of a continuous network which is used to improve the closed-loop behavior; $G_f(s)$ is the transfer function of the fixed elements. In Fig. 9-12B, $G_c^*(s)$ is the transfer function of a digital network (digital program) used to improve system performance. For the case of continuous compensation, we have

$$\frac{E^*(s)}{R^*(s)} = \frac{1}{1 + [G_c(s)G_f(s)]^*} \quad (9-44)$$

$$\frac{C(s)}{R^*(s)} = \frac{G_c(s)G_f(s)}{1 + [G_c(s)G_f(s)]^*} \quad (9-45)$$

For digital compensation, we have

$$\frac{E^*(s)}{R^*(s)} = \frac{1}{1 + G_f^*(s)G_c^*(s)} \quad (9-46)$$

$$\frac{C(s)}{R^*(s)} = \frac{G_f(s)G_c^*(s)}{1 + G_f^*(s)G_c^*(s)} \quad (9-47)$$

Compensation with a digital network is a simpler analytical problem than continuous compensation because the effect of the compensation can be varied without altering the fixed-element contribution to the open-loop response $G_f^*(s)G_c^*(s)$. For continuous compensation, the open-loop response is $[G_c(s)G_f(s)]^*$, and altering the compensation will alter the contribution of the fixed elements to the open-loop response. Conventional continuous system techniques can be used for synthesizing digital programs to compensate sampled-data systems; but if continuous compensation is desired, a trial-and-error analytical procedure is necessary.

9-5 PERFORMANCE EVALUATION

The determination of the time response of a sampled-data system can be carried out in closed form by the use of Eqs. (9-38) and (9-39). A simple numerical procedure for determining the values of the output at the sampling instants can be obtained by expanding the z transform of the output in a power series in z . Since z corresponds to a time delay of one sampling instant, the coefficients of the power-series expansion of a z transform are the values of the corresponding time function at the sampling instants, as can be seen from an examination of Eq. (9-10). The expansion is easily performed by dividing the numerator by the denominator since the z transform is a rational function.

Example.

Assume that the z transform of the output is given by:

$$C^*(z) = \frac{0.186(z^2 + 1.392)}{(1 - z)[0.554z^2 - 1.108z + 1]}$$

Dividing the numerator by the denominator, the power series expansion of $C^*(z)$ is found to be

$$\begin{aligned} C^*(z) = & (.26)z + (.76)z^2 + (1.17)z^3 \\ & + (1.36)z^4 + (1.33)z^5 + (1.20)z^6 \\ & + (1.07)z^7 + (.99)z^8 + (.96)z^9 + \dots \end{aligned}$$

The coefficients of this expansion are the sampled values of $c(t)$, and the instant of occurrence is determined from the power of z in the appropriate term. The function is plotted in Fig. 9-13.

An alternate method for finding the time response of a system is based on a difference-equation representation. Assume that

$$\frac{C^*(s)}{R^*(s)} = \frac{a_0 + a_1e^{-sT} + a_2e^{-2sT} + \dots}{b_0 + b_1e^{-sT} + b_2e^{-2sT} + \dots} \quad (9-48)$$

SAMPLED-DATA SYSTEMS

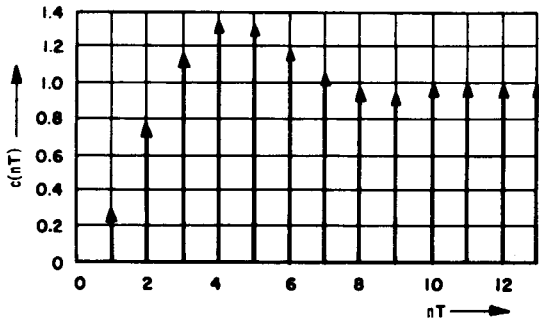


Fig. 9-13 Sampled-time function.

Cross-multiplying, inverse transforming, and solving for $c(nT)$, we obtain

$$c(nT) = \frac{1}{b_0} [a_0 r(nT) + a_1 r[(n-1)T] + \dots - b_1 c[(n-1)T] - b_2 c[(n-2)T] - \dots] \quad (9-49)$$

This is a general recurrence formula which enables one to calculate the present value $c(nT)$ of the output in terms of a weighted sum of the present and past values of the input $r(t)$ and the past values of the output. The calculation is best carried out in tabular form.

Example. Assume that the closed-loop transfer function of a sampled-data servomechanism is

$$\frac{C^*(z)}{R^*(z)} = \frac{1.5z}{1 + 0.5z}$$

If $r(t)$ is a unit step, the tabular evaluation of $c(nT)$ can be carried out as follows. Cross-multiplying, we get

$$C^* + 0.5 e^{-T} C^* = 1.5 e^{-T} R^*$$

Inverse transforming, we obtain

$$c(nT) + 0.5 c[(n-1)T] = 1.5 r[(n-1)T]$$

or

$$c(nT) = 1.5 r[(n-1)T] - 0.5 c[(n-1)T]$$

when $r(nT)$ is a unit step, and where the system has been at rest so that $c(-T)$ is zero, the calculation of $c(nT)$ can be carried out in tabular form as follows:

| n | $1.5r[(n-1)T]$ | $-0.5c[(n-1)T]$ | $c(nT)$ |
|-----|----------------|-----------------|---------|
| 0 | 0 | 0 | |
| 1 | 1.5 | 0 | 0 |
| 2 | 1.5 | -0.75 | 1.5 |
| 3 | 1.5 | -0.375 | 0.75 |
| 4 | 1.5 | -0.562 | 1.125 |
| 5 | 1.5 | -0.469 | 0.938 |
| 6 | 1.5 | -0.516 | 1.031 |
| 7 | 1.5 | -0.492 | 0.984 |
| 8 | 1.5 | -0.504 | 1.008 |
| 9 | 1.5 | -0.498 | 0.996 |
| | | | 1.002 |

The values in the last column are plotted in Fig. 9-14. In general, it is necessary to know the value of $c(nT)$ for values of n corresponding to time prior to the beginning of the transient. If the system is of k th order, it is necessary to know $c(nT)$ for k prior to the samples. This corresponds to the need for knowing the initial conditions in any transient problem.

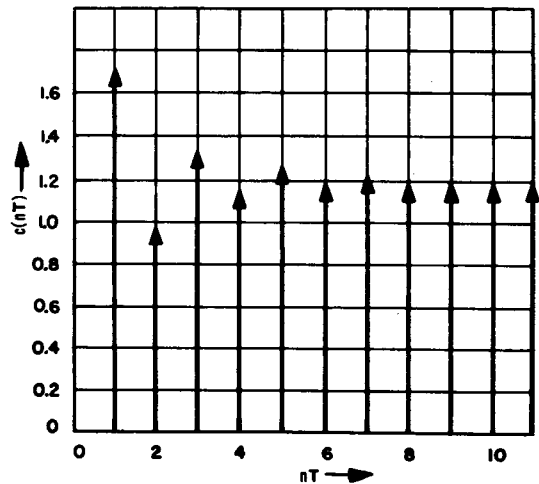


Fig. 9-14 Step response of sampled-data system.

THEORY

Since the z transform does not give the value of the output between sampling instants, Lago and Truxal⁽¹⁰⁾ have suggested a method which enables one to determine the output at submultiples of the sampling period T . A fictitious impulse modulator is placed immediately after the actual one (Fig. 9-15). The fictitious sampler has a sampling period which is an integral submultiple of T ; i.e.,

$$T' = \frac{T}{n} \quad (9-50)$$

As a result, we can write

$$C^*(z) = \frac{R^*(z^n)}{1 + G^*(z^n)} G'^*(z) \quad (9-51)$$

where $G'^*(z)$ and $C'^*(z)$ are z transforms of $g(t)$ and $c(t)$ with respect to the period T' , and $R^*(z^n)$ and $G^*(z^n)$ are z transforms of $r(t)$ and $g(t)$ with respect to the period T but with z replaced by z^n .

An extension of the method above is given by Barker.⁽²⁴⁾ In this extended method, the output at *any* time between sampling instants can be found. If we refer to Fig. 9-16, it can be seen that the artificial delay produces a signal $c'(t)$ which one can sample in order to observe the values which will occur between the values of $c^*(t)$. We find that

$$C'^*(z, m) = \frac{G^*(z, m)}{1 + G^*(z)} R^*(z) \quad (9-52)$$

where $G^*(z, m)$ is a modified z transform. $G^*(z, m)$ is evaluated by assuming that $g(t)$ is sampled at $t = (n + m - 1)T$ instead of at $t = nT$. A brief table of modified z transforms is listed in Table 9-2. A more extensive table is given by Barker.⁽²⁴⁾ The use of Eq. (9-52) enables one to scan the output by varying m between zero and unity. Thus, the variation of the output between sampling instants is observed, and a study of the ripple can be made.

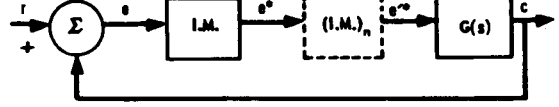


Fig. 9-15 Determination of $c(t)$ between sampling instants by sampling at $n \Omega$ rad/sec.

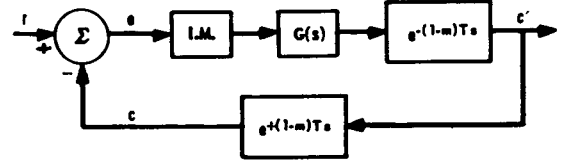


Fig. 9-16 Determination of $c(t)$ between sampling instants through the use of an artificial delay.

The *error coefficients* of a sampled-data system may be obtained from the expression

$$e_n = \frac{1}{n!} \left\{ \frac{d^n}{ds^n} \left[\frac{E^*(s)}{R^*(s)} \right] \right\}_{s=0} \quad (9-53)$$

Since $E^*(s)/R^*(s)$ is a rational function in z , where $z = e^{-sT}$, we can expand it in a Taylor series (in z) about the point $z = 1$ ($s = 0$). Then, by using the infinite series expansion of e^{-sT} and rearranging terms, we can easily obtain the Taylor series expansion of $E^*(s)/R^*(s)$ in terms of s at $s = 0$:

$$\begin{aligned} \frac{E^*(z)}{R^*(z)} &= a_0 + a_1(z - 1) + a_2(z - 1)^2 \\ &\quad + a_3(z - 1)^3 + \dots \end{aligned} \quad (9-54)$$

SAMPLED-DATA SYSTEMS

or

$$\frac{E^*(s)}{R^*(s)} = a_0 - T(a_1)s + T^2(a_2 + \frac{a_1}{2})s^2 - T^3(a_3 + a_2 + \frac{a_1}{3})s^3 + \dots \quad (9-55)$$

The techniques described in this section can be used to obtain the transient response, the ripple, and the error coefficients of a sampled-data system. Using trial-and-error procedures and the conventional continuous-system design techniques, the evaluation of a given system in terms of a set of performance specifications is a straightforward matter.

TABLE 9-2 MODIFIED z TRANSFORMS

| $F(s)$ | $F^*(z, m)$ |
|-----------------|--------------------------------|
| 1 | 0 |
| $\frac{1}{s}$ | $\frac{z}{1-z}$ |
| $\frac{1}{s^2}$ | $\frac{Tz^2 + mz}{(1-z)^2}$ |
| $\frac{1}{s+a}$ | $\frac{ze^{-amT}}{1-ze^{-aT}}$ |

BIBLIOGRAPHY

- 1 W. K. Linvill, "Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation", *Trans. AIEE*, Vol. 70, Part II, pp. #1779-1788, 1951.
- 2 K. S. Miller and R. J. Schwarz, "Analysis of a Sampling Servomechanism", *J. Appl. Phys.*, Vol. 21, pp. #290-294, 1950.
- 3 J. R. Ragazzini and L. A. Zadeh, "The Analysis of Sampled-Data Systems", *Trans. AIEE*, Vol. 71, Part II, pp. #225-234, 1952.
- 4 R. G. Brown and G. J. Murphy, "An Approximate Transfer Function for the Analysis and Design of Pulsed Servos", *Trans. AIEE*, Vol. 71, Part II, pp. #435-440, 1952.
- 5 E. I. Jury, "Analysis and Synthesis of Sampled-Data Control Systems", *Trans. AIEE*, Vol. 73, Part I, pp. #332-346, 1954.
- 6 W. K. Linvill and J. M. Salzer, "Analysis of Control Systems Involving Digital Computers", *Proc. IRE*, Vol. 41, pp. #901-906, July, 1953.
- 7 A. R. Bergen and J. R. Ragazzini, "Sampled-Data Processing Techniques for Feedback Control Systems", *Trans. AIEE*, Vol. 73, Part II, pp. #236-247, 1954.
- 8 C. K. Chow, "Contact Servomechanisms Employing Sampled Data", *Trans. AIEE*, Vol. 73, Part II, pp. #51-64, 1954.
- 9 G. W. Johnson and D. P. Lindorff, "Transient Analysis of Sampled-Data Control Systems", *Trans. AIEE*, Vol. 73, Part II, pp. #147-153, 1954.
- 10 G. V. Lago and J. G. Truxal, "The Design of Sampled-Data Feedback Systems", *Trans. AIEE*, Vol. 73, Part II, pp. #247-253, 1954.

- 11 G. V. Lago, "Additions to z-Transformation Theory for Sampled-Data Systems", *Trans. AIEE*, Vol. 73, Part II, pp. #403-408, 1954.
- 12 E. I. Jury, "The Effect of Pole and Zero Locations on the Transient Response of Sampled-Data Systems", *Trans. AIEE*, Vol. 74, Part II, pp. #41-48, 1955.
- 13 J. Sklansky and J. R. Ragazzini, "Analysis of Errors in Sampled-Data Systems", *Trans. AIEE*, Vol. 74, Part II, pp. #65-71, 1955.
- 14 G. W. Johnson, D. P. Lindorff, and C. G. A. Nordling, "Extension of Continuous-Data System Design Techniques to Sampled-Data Control Systems", *Trans. AIEE*, Vol. 74, Part II, pp. #252-263, 1955.
- 15 E. I. Jury, "Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems", *Trans. AIEE*, Vol. 74, Part II, pp. #427-435, 1955.
- 16 K. K. Maitra and P. E. Sarachik, "Digital Compensation of Continuous-Data Feedback Control Systems", *Trans. AIEE*, Vol. 75, Part II, pp. #107-116, 1956.
- 17 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms*, MIT Radiation Laboratory Series, Vol. 25, pp. #231-261, McGraw-Hill Book Co., Inc., New York, N. Y., 1947.
- 18 H. S. Tsien, *Engineering Cybernetics*, pp. #83-93, McGraw-Hill Book Co., Inc., New York, N. Y., 1954.
- 19 J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. #500-558, McGraw-Hill Book Co., Inc., New York, N. Y., 1955.
- 20 B. M. Brown, "Application of Finite Difference Operators to Linear Systems", in *Automatic and Manual Control*, edited by A. Tustin, pp. #409-418, Butterworths Scientific Publications, London, England, 1953.
- 21 C. Holt Smith, D. F. Lawden, and A. E. Bailey, "Characteristics of Sampling Servo Systems", in *Automatic and Manual Control*, edited by A. Tustin, pp. #377-404, Butterworths Scientific Publications, London, England, 1952.
- 22 R. C. Oldenbourg, "Deviation Dependent Step-by-Step Control as Means to Achieve Optimum Control for Plants with Large Distance-Velocity Lag", in *Automatic and Manual Control*, Proceedings of Cranfield Conference, 1951, edited by A. Tustin, pp. #435-447, Butterworths Scientific Publications, London, England, 1952.
- 23 H. Sartorius, "Deviation Dependent Step-by-Step Control Systems and Their Stability", in *Automatic and Manual Control*, Proceedings of Cranfield Conference, 1951, edited by A. Tustin, pp. #421-434, Butterworths Scientific Publications, London, England, 1952.
- 24 R. H. Barker, "The Pulse Transfer Function and Its Application to Sampling Servosystems", *Proc. IEE* (London), Vol. 99, Part IV, 1952.
- 25 W. M. Stone, "A List of Generalized Laplace Transforms", *Journal of Science*, Iowa State College, Ames, Iowa, Vol. 22, pp. #215-225, April, 1948.
- 26 M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, Vol. I, pp. #354-356, John Wiley & Sons, Inc., New York, N. Y., 1942.
- 27 D. F. Lawden, "A General Theory of Sampling Servosystems", *Proc. IEE* (London), Vol. 98, Part IV, pp. #31-36, October, 1951.
- 28 I. N. Sneddon, *Fourier Transforms*, McGraw-Hill Book Co., Inc., New York, N. Y., 1951.

CHAPTER 10

NONLINEAR SYSTEMS***10-1 INTRODUCTION**

All of the techniques of system analysis discussed in previous chapters are restricted in their application to linear systems. This restriction imposes two limitations on design. First, components must be of high quality if they are to operate in a linear manner when amplitudes and frequencies of signals vary widely. Second, the linearity restriction limits the realizable system characteristics, the types of systems, and the tasks that can be accomplished.

Nonlinearities are generally of two types:

incidental and intentional. Incidental nonlinearities are secondary effects which limit performance in otherwise linear systems. Examples of phenomena that introduce incidental nonlinearities include backlash, saturation, dead zone, hysteresis, and coulomb friction. On the other hand, intentional nonlinearities are those introduced purposely to improve the characteristics of systems or to alter them in specified ways. The contactor (on-off or relay) servo is the most extreme example of such an intentionally nonlinear system.

10-2 DESCRIBING FUNCTION PROCEDURES (7,8,12,13,15,18,19,21,22,31,35,36,42,50,51,52)

One problem to be analyzed in an investigation of nonlinear system behavior is that concerned with the question of stability. A method of studying this problem utilizes the describing-function procedure. The application of the *describing-function* procedure enables the designer to predict whether or not a closed-loop system containing a nonlinear element will be stable. A system is said to be stable if, after a sudden input or disturbance, it eventually comes to rest. In some systems the existence of a stable oscillation is acceptable provided the amplitude of the oscillation is small. A typical case is a relay control system where a small amplitude oscillation may be acceptable if the cost of eliminating the

oscillation is too high. A system is said to be unstable if a finite input or disturbance to the system results in an output oscillation that tends to grow without bound.

The describing-function method is based on three assumptions:

(a) There is only one nonlinear element in the system. If there are more than one, that part of the system including all nonlinearities is treated as a single nonlinear component.

(b) The characteristics of the nonlinear element are independent of time. They depend only on the present value and past history of the input to the element.

(c) If the input to the nonlinear element is sinusoidal, only the fundamental sinusoidal component of the output of the element contributes to the input of this element.

*By L. A. Gould

NONLINEAR SYSTEMS

The last assumption is the heart of the method. It is applicable when the amplitude of the harmonics generated by the nonlinearity decreases and when the elements that follow the nonlinearity have low-pass characteristics.

Referring to Fig. 10-1, the method of analysis is as follows. The input $x(t)$ to the nonlinear element N is assumed to be sinusoidal with amplitude X and frequency ω . The output $y(t)$ will be periodic (but nonsinusoidal) with the same frequency ω . A Fourier analysis of the output waveform is made and all frequencies except the fundamental are ignored. The amplitude Y_1 of the fundamental will, in general, be a function of X and ω , as will be the phase angle of the fundamental relative to the input. The describing function of the nonlinear element is defined as a complex number whose magnitude is Y_1/X , the ratio of the amplitude of the fundamental component of the output to the amplitude of the input, and whose angle is the phase angle of the funda-

mental component of the output relative to the phase angle of the input. The describing function is usually denoted by $N(X, \omega)$. Symbolically, if

$$x(t) = X \sin \omega t \quad (10-1)$$

then

$$y(t) = Y_1 \sin (\omega t + \phi_1) + Y_2 \sin (\omega t + \phi_2) + Y_3 \sin (\omega t + \phi_3) + \dots \quad (10-2)$$

and

$$\left| N(X, \omega) \right| = \frac{Y_1}{X} \quad (10-3)$$

$$\angle N(X, \omega) = \phi_1 \quad (10-4)$$

The describing functions of several important nonlinearities will now be presented. Figure 10-2 shows the input-output characteristic of a *contactor* with inactive zone Δ (dead zone) and hysteresis h . The amplitude and phase curves of the describing function N of

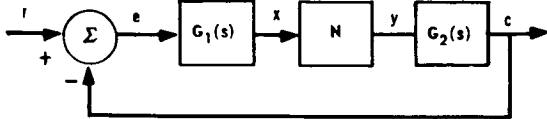


Fig. 10-1 Nonlinear feedback control system.

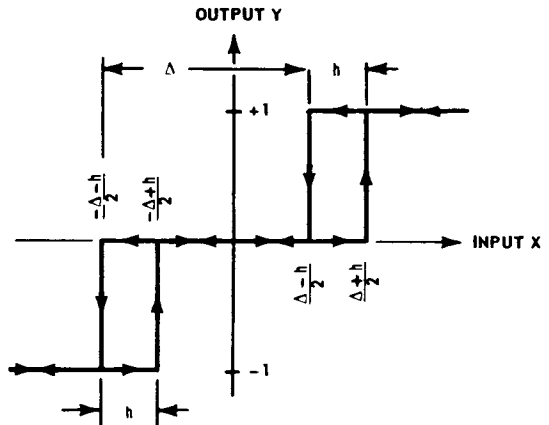


Fig. 10-2 Dimensionless representation of contactor characteristics (case involving both inactive zone and hysteresis).

Adapted by permission from *Transactions of the AIEE*, Volume 69, Part I, 1950, from article entitled 'A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms', by R. J. Kochenburger.

THEORY

this contactor as functions of the input amplitude X appear in Fig. 10-3. Figure 10-4 shows the input-output characteristics of a *nonlinear element containing both dead zone D and saturation S* . No phase shift is associated with this element since, in general, phase shift will not occur for a single-valued nonlinearity. The describing function for no saturation ($S \rightarrow \infty$) is presented in Fig. 10-5. The describing function for no dead zone ($D = 0$) is presented in Fig. 10-6. Describing functions for various combinations of dead zone and saturation

appear in Fig. 10-7. Figure 10-8 shows the input-output characteristic of a nonlinear element characterized by *hysteresis* (backlash or free play). The describing function for this nonlinearity is presented in Fig. 10-9. Other describing functions for more complex nonlinearities can be derived for the particular case being considered. Additional describing functions are given in the literature.^(18,19,22)

The procedure for using the describing function to predict the nature of the stability of a nonlinear system follows. Referring to

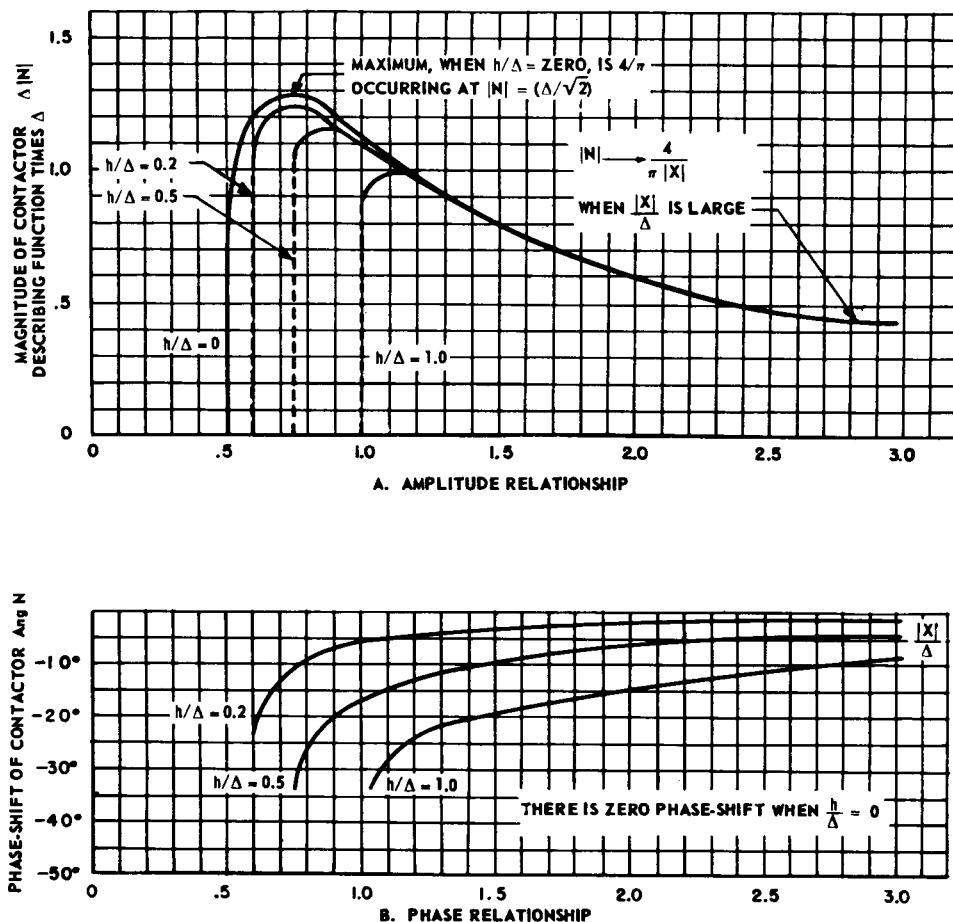


Fig. 10-3 Plot of the describing function N (simple contactor with hysteresis ratio h/Δ).

Adapted by permission from *Transactions of the AIEE*, Volume 69, Part I, 1950, from article entitled 'A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms', by R. J. Kochenburger.

NONLINEAR SYSTEMS

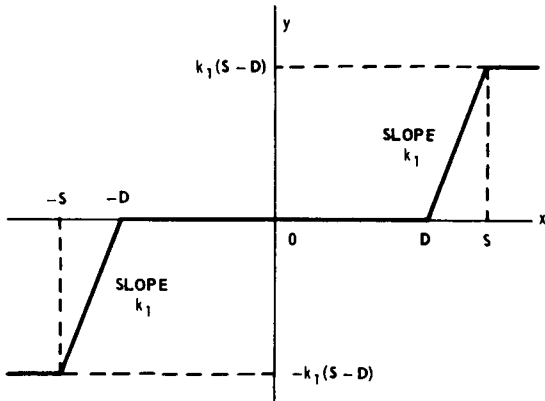


Fig. 10-4. Nonlinear characteristic with dead zone and saturation.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

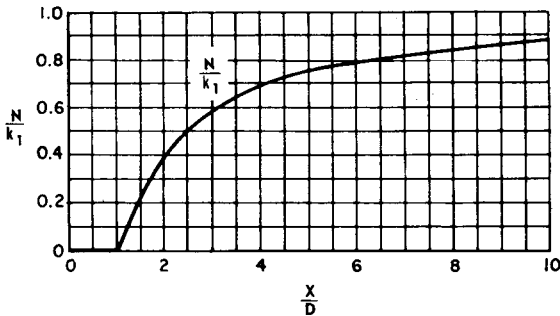


Fig. 10-5 Describing function for dead zone.

Adapted by permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

Fig. 10-10, the linear and nonlinear portions of the system are separated into two parts; the describing function N applies to one part, and the response of the linear elements G to the other. The *gain-phase plane* is employed for plotting the negative reciprocal ($-1/N$) of the describing function. The response $G(j\omega)$ of the linear elements is also plotted on

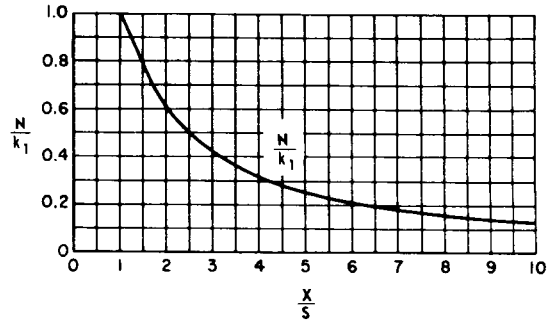


Fig. 10-6 Describing function for saturation.

Adapted by permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

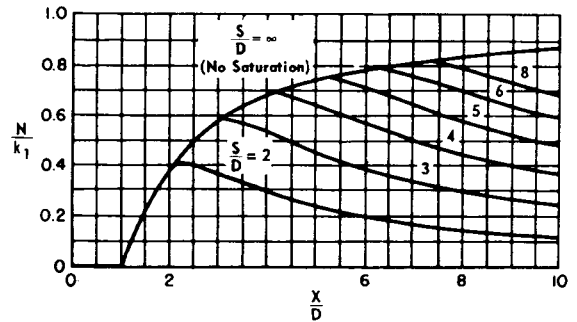


Fig. 10-7 Describing function for saturation and dead zone.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

the same plane. If the $-1/N$ locus and the $G(j\omega)$ locus do not intersect, the system is stable and does not oscillate. If the $-1/N$ locus and the $G(j\omega)$ locus do intersect (two types of intersections can occur), the system may or may not be oscillatory. The describing function for a contactor with hysteresis and dead zone is sketched in Fig. 10-11 wherein the types of intersections of the $-1/N$ locus with a $G(j\omega)$ locus are shown. The parameter along the $-1/N$ locus is the amplitude X of the assumed sinusoidal input to the nonlinear

THEORY

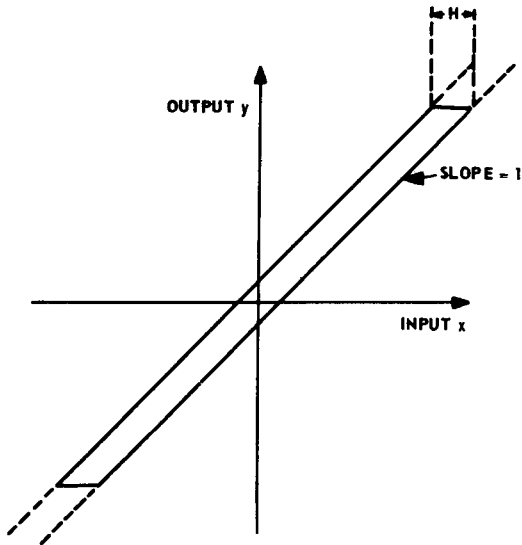


Fig. 10-8 Hysteresis nonlinearity.

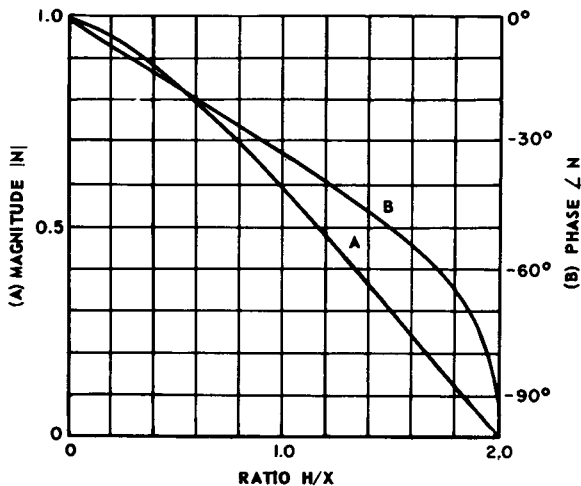


Fig. 10-9 Describing function for hysteresis-type nonlinear element.

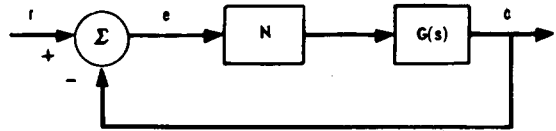


Fig. 10-10 Simplified nonlinear system.

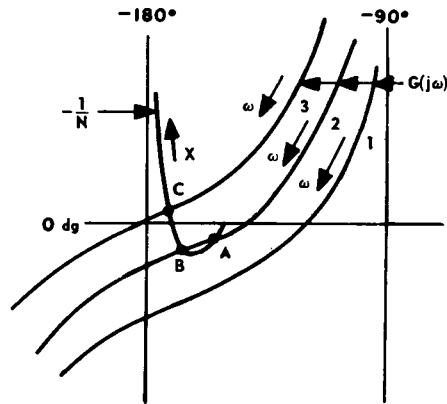


Fig. 10-11 Stability determination with describing function.

element. Three cases illustrating the types of intersection of the $G(j\omega)$ locus with the $-1/N$ locus are shown in Fig. 10-11.

Case 1. The $G(j\omega)$ locus does not intersect the $-1/N$ locus. The system is stable and no oscillation occurs.

Case 2. The $G(j\omega)$ locus intersects the $-1/N$ locus at two points A and B. Point A is called a *divergent equilibrium* point since sustained oscillations cannot be maintained at the frequency ω_A and amplitude X_A associated with A. The existence of a divergent equilibrium can be determined by treating the $-1/N$ locus as one treats the $-1 + j0$ point in the study of the stability of linear systems. If the amplitude X associated with point A decreases slightly, the $G(j\omega)$ locus will be located in a stable position with respect to the $-1/N$ point, and oscillations will tend to die out. If the amplitude X tends to increase from X_A , the $G(j\omega)$ locus encloses the $-1/N$ point

Adapted by permission from *Transactions of the AIEE*, Volume 72, Part II, 1953, from article entitled 'Describing Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities', by H. D. Grief.

(corresponding to instability), and the amplitude of oscillation will tend to increase. Thus, oscillation cannot be maintained at A . However, point B is a point of *convergent equilibrium*, as can be determined by letting the amplitude X both increase and decrease relative to X_B . In each case, the tendency will be for the amplitude of the oscillation to head back to point B . Thus, the convergent equilibrium point B determines an amplitude X_B (read off the $-1/N$ locus) and a frequency ω_B [read off the $G(j\omega)$ locus] at which a sustained oscillation occurs.

Case 3. The $G(j\omega)$ locus intersects the $-1/N$ locus at point C . This is a *convergent equilibrium* point, as can be determined by letting the amplitude X increase or decrease relative to the amplitude X_c associated with the intersection C . In each case, the amplitude of oscillation will tend to return to X_c .

The describing-function method thus predicts the stability of nonlinear systems, as described above. If intersections occur between the $-1/N$ locus and the $G(j\omega)$ locus, the amplitude and frequency of convergent oscillations can be predicted to an accuracy that is determined by the assumptions inherent in the method. Techniques for estimating the accuracy of the results are given in references 15 and 52.

The describing-function procedure breaks down if the two loci ($-1/N$ and G) approach each other without intersecting, are tangent, or intersect at a small angle. In these situations one cannot be certain as to whether or not oscillations exist.

In general, the accuracy of the describing-function method increases as the cutoff rate of the linear element (*following* the nonlinear element) increases. The accuracy may decrease if the linear element exhibits a sharp resonance.

The describing-function procedure is useful in predicting the closed-loop frequency response of a system containing an incidental nonlinearity when no oscillation can occur. Thereby, peculiarities in measured characteristics can be explained, and quantitative

estimates of nonlinear effects can be made. By treating the $-1/N$ locus as the equivalent of the $-1 + j0$ point of conventional linear analysis, the degree of stability of a system containing a nonlinear element may be estimated. Instead of aligning the $(-180^\circ, 0 \text{ dg})$ point of the Nichols chart with the $(-180^\circ, 0 \text{ dg})$ point of the $G(j\omega)$ locus in order to determine M_p , as is done when a linear element is present, the $(-180^\circ, 0 \text{ dg})$ point of the Nichols chart is aligned with a point chosen on the $-1/N$ locus of the nonlinear element for a given amplitude X of the input to the element, when a nonlinear element is present. The tangency of the G locus (plotted on the same coordinates as the $-1/N$ locus) to an M contour of the Nichols chart will then be an indication of the degree of stability associated with the chosen amplitude X . Moving the $(-180^\circ, 0 \text{ dg})$ point of the Nichols chart along the $-1/N$ locus is equivalent to changing the amplitude X of the input to the nonlinear element. By this means, the variation of the degree of stability (as measured by M_p) can be determined as a function of the amplitude of the input to the nonlinear element. The relation between the amplitude X (input to nonlinear element) and the reference-input amplitude R can be determined for each M_p value from the following relations (see Fig. 10-1).

$$|R| = \frac{|N| |G_2(j\omega_p)|}{M_p} |X| \quad (10-5)$$

where ω_p is the frequency associated with the point of M_p tangency for each value of X along the $-1/N$ locus.

These methods can be extended to determine the entire frequency response of a system by noting the intersections of other M contours with the G locus at each value of X and using the following relations to determine the input amplitude (or amplitudes) R associated with each value of X :

$$\left| \frac{C(j\omega)}{R(j\omega)} \right| = \left| \frac{G_1(j\omega) N G_2(j\omega)}{1 + G_1(j\omega) N G_2(j\omega)} \right| \quad (10-6)$$

$$\left| \frac{X(j\omega)}{R(j\omega)} \right| = \frac{1}{|N G_2(j\omega)|} \left| \frac{C(j\omega)}{R(j\omega)} \right| \quad (10-7)$$

10-3 PHASE-PLANE PROCEDURES (2,3,4,5,6,23,24,25,26,27,30,34,39,40,41,44,46,47,48,51,52,53)

The main limitation of the describing-function procedure is that it cannot predict the response of stable nonlinear systems to inputs that are not sinusoidal. In contrast, the *phase-plane method* is an attempt to describe the response of nonlinear systems to specific transient inputs. In this method, attention is focused on the differential equations that describe the system, and the behavior of the system is studied by plotting velocity versus displacement with time as a parameter. This velocity-displacement plane is called the *phase plane*. Only second-order systems can be handled in the phase plane although attempts have been made to treat higher-order systems by a phase-space representation. (34,40,46,47,48,53)

Phase-plane analysis is concerned with the characteristics of the differential equation

$$\ddot{x} + a(x, \dot{x})\dot{x} + b(x, \dot{x})x = 0 \quad (10-8)$$

The phase-plane portrait of the system is a plot of the velocity \dot{x} as a function of the displacement x , the plot being a family of curves depending on the initial conditions $x(0)$ and $\dot{x}(0)$. Once the initial conditions have been specified, the behavior of the system is determined completely by the curve in the phase plane corresponding to the given initial conditions. Thus, the phase-plane approach is most useful in determining the response of a system to a step input. Since a step input does not always occur in practice, the application of the phase-plane technique is severely restricted when the response to other types of inputs is sought.

As an example, consider a second-order *linear* system whose characteristic equation is

$$\ddot{x} + 2\zeta\omega_n \dot{x} + \omega_n^2 x = 0 \quad (10-9)$$

If the velocity \dot{x} is treated as a new variable y , then, by eliminating the dependence of the above equation on time, there results

$$\frac{dy}{dx} + 2\zeta\omega_n + \omega_n^2 \frac{x}{y} = 0 \quad (10-10)$$

Equation (10-10) is a first-order equation for y as a function of x and has a family of solutions depending on the initial values $y(0)$ and $x(0)$. Each solution is called a *phase trajectory*, and the totality of solutions is the phase-plane portrait of the system. The phase trajectories for Eq. (10-10) are shown in Fig. 10-15 for $\zeta = 0.5$. In this figure, if the initial conditions correspond to the point A_0 , then the motion of the system is completely described by the trajectory $A_0A_1A_2A_3A_4A_5A_6$ with time increasing in the direction of the arrows.

In a more general case, it may be very difficult to solve the equation that describes the trajectories. A graphical procedure involving the determination of the *isoclines* (lines of equal slope) is then possible.⁽⁴⁾ Referring to Eq. (10-8), the slope of the phase trajectories is found to be

$$\frac{dy}{dx} = -a(x, y) - b(x, y) \frac{x}{y} \quad (10-11)$$

By setting the right side of this equation equal to a constant, a curve connecting points of equal slope is determined. The isoclines thus obtained are plotted in the phase plane, and the slopes of the various phase trajectories can be drawn directly on the isoclines. If a large number of isoclines are drawn, the phase trajectories can be accurately determined.

Once the phase portrait of a system has been constructed, the behavior of the system can be investigated. If the response of the system for a given set of initial conditions is sought, the corresponding phase trajectory determines the response. The variation of time t along the trajectory can be ascertained from the relation

$$t = \int \frac{1}{y} dx \quad (10-12)$$

The nature of the stability of the system can be determined by an investigation of the *singular points* of the system. If the behavior

THEORY

of a second-order system can be described by the two first-order equations

$$\dot{x} = P(x, y) \quad (10-13)$$

$$\dot{y} = Q(x, y) \quad (10-14)$$

the points at which $\dot{x} = 0$ and $\dot{y} = 0$ are called the *singular points* of the system and represent *equilibrium states* of the system. If the trajectories approach a singular point, the system is stable; whereas, if they diverge from the singular point, the system may be unstable. To investigate the nature of the equilibrium at a singular point, a Taylor series expansion of the functions $P(x, y)$ and $Q(x, y)$ is made about the point, and all but

the first-order terms in expansion are neglected. Thus, the singular points are determined from the solutions of the equations

$$P(x, y) = 0 \quad (10-15)$$

$$Q(x, y) = 0 \quad (10-16)$$

The linearized forms of Eqs. (10-13) and (10-14) at the singular point $x = a$ and $y = b$ become

$$\dot{x} = a_1(x - a) + a_2(y - b) \quad (10-17)$$

$$\dot{y} = b_1(x - a) + b_2(y - b) \quad (10-18)$$

where a_1 , a_2 , b_1 , and b_2 are coefficients of the expansion. Six types of singular points can occur:

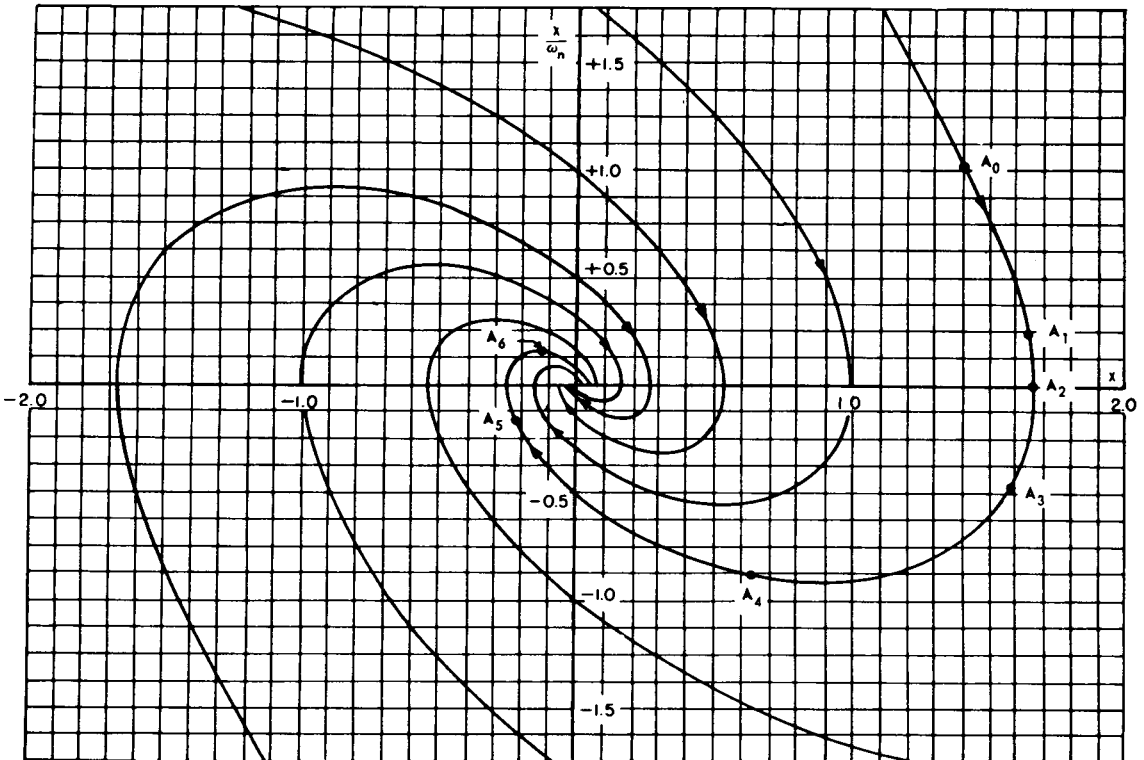


Fig. 10-15 Phase portrait of linear second-order system with $\zeta = 0.5$.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

(a) *Stable node* (Fig. 10-16)

$$(a_1 + b_2) < 0 \quad (10-19)$$

$$(a_1 + b_2)^2 > 4(a_1b_2 - a_2b_1) \quad (10-20)$$

(b) *Unstable node* (Fig. 10-17)

$$(a_1 + b_2) > 0 \quad (10-21)$$

$$(a_1 + b_2)^2 > 4(a_1b_2 - a_2b_1) \quad (10-22)$$

(c) *Stable focus* (Fig. 10-18)

$$(a_1 + b_2) < 0 \quad (10-23)$$

$$(a_1 + b_2)^2 < 4(a_1b_2 - a_2b_1) \quad (10-24)$$

(d) *Unstable focus* (Fig. 10-19)

$$(a_1 + b_2) > 0 \quad (10-25)$$

$$(a_1 + b_2)^2 < 4(a_1b_2 - a_2b_1) \quad (10-26)$$

(e) *Center* (Fig. 10-20)

$$(a_1 + b_2) = 0 \quad (10-27)$$

$$(a_1b_2 - a_2b_1) > 0 \quad (10-28)$$

(f) *Saddle point* (Fig. 10-21)

$$(a_1 + b_2) = 0 \quad (10-29)$$

$$(a_1b_2 - a_2b_1) < 0 \quad (10-30)$$

The relations among the various singular points and the Taylor series coefficients given by Eqs. (10-19) through (10-30) are summarized in Fig. 10-22.

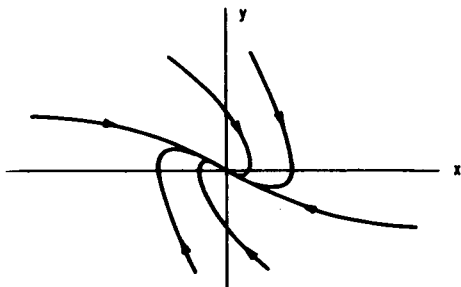


Fig. 10-16 Portrait in the vicinity of a stable node.

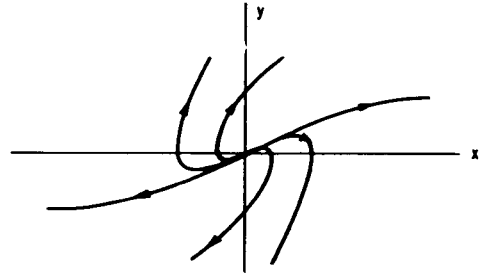


Fig. 10-17 Portrait in the vicinity of an unstable node.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

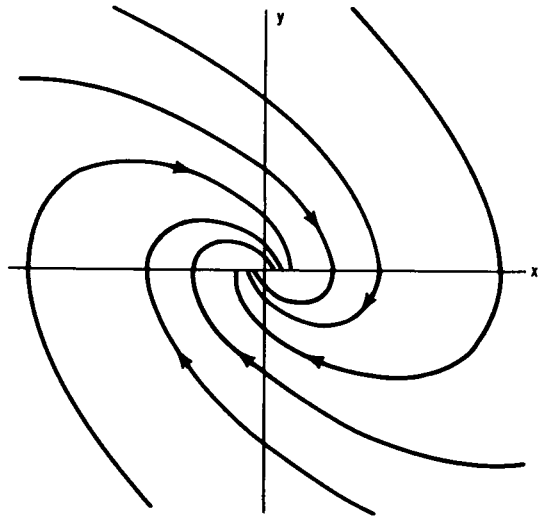


Fig. 10-18 Portrait in the vicinity of a stable focus.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

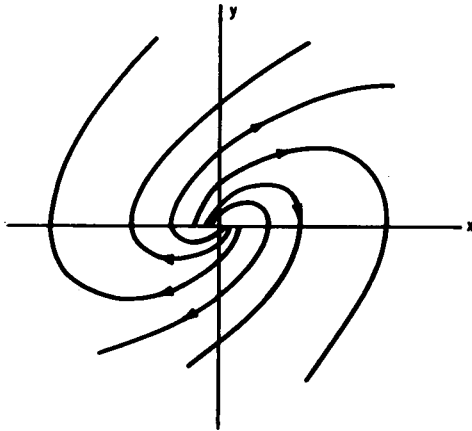


Fig. 10-19 Portrait in the vicinity of an unstable focus.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

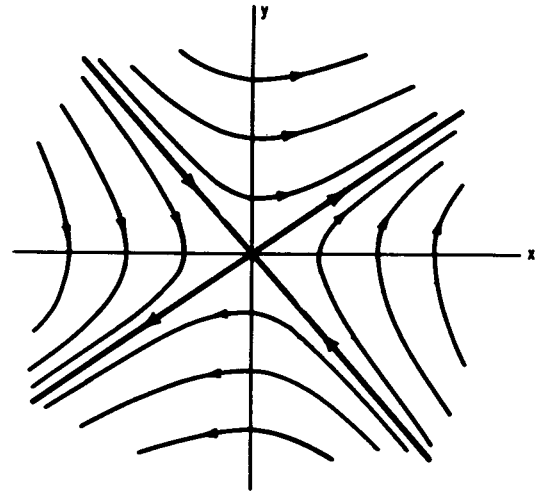


Fig. 10-21 Portrait in the neighborhood of a saddle point.

Adapted by permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

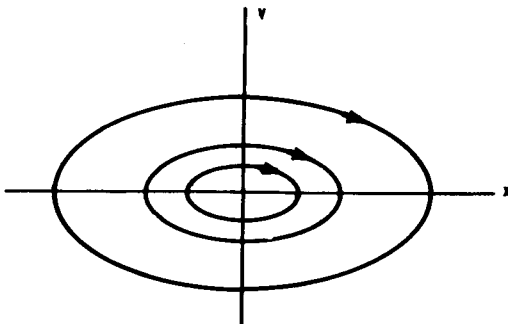


Fig. 10-20 Portrait in the vicinity of a center.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

In the case of feedback control systems, the problem is simplified because Eq. (10-13) can be replaced by

$$\dot{x} = y \quad (10-31)$$

and the Taylor series coefficients a_1 and a_2 become

$$a_1 = 0 \quad (10-32)$$

$$a_2 = 1 \quad (10-33)$$

In addition to the determination of the singular points, a complete description of the stability of a system in the phase plane requires a determination of the *limit cycles* of the system. A *limit cycle* is an isolated closed path in the phase portrait which corresponds to a system oscillation of fixed amplitude and period. A limit cycle is stable or unstable depending upon whether the paths in the neighborhood converge toward the limit cycle or diverge away from it. Thus, there arise two general types of self-excitation of non-linear systems. *Soft excitation* occurs when a limit cycle encloses an unstable singular point (Fig. 10-23); *hard excitation* occurs when a limit cycle encloses a stable limit cycle or a stable singular point (Fig. 10-24). There is no definite method available for determining the limit cycles of a system or even if a limit cycle exists. The only approach is to determine the convergent and divergent properties of the phase trajectories. Thus, if all trajectories are converging outside a circle C_1 (centered at the origin) and diverging inside a smaller circle C_2 (centered at the origin), then a stable limit cycle must exist between

NONLINEAR SYSTEMS

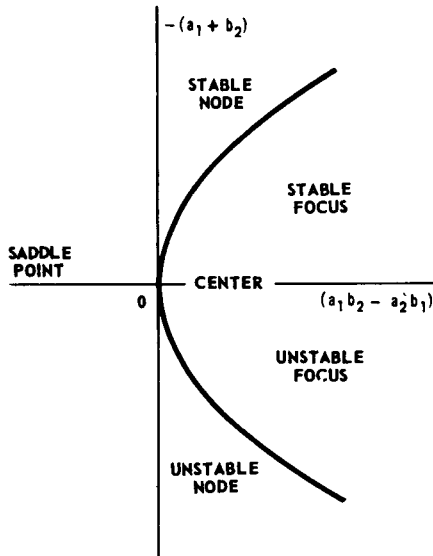


Fig. 10-22 Types of singularities.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

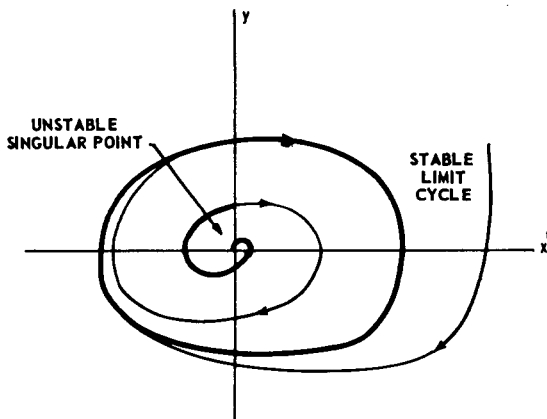


Fig. 10-23 Portrait with soft self-excitation.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

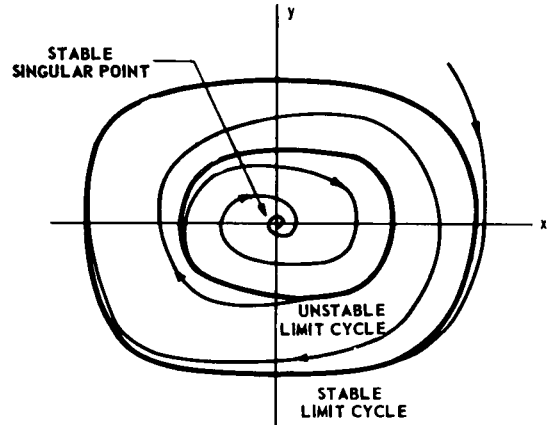


Fig. 10-24 Portrait with hard self-excitation.

By permission from *Automatic Feedback Control System Synthesis*, by J. G. Truxal, Copyright, 1955, McGraw-Hill Book Company, Inc.

the two circles. In particular, an examination of the time rate of change of the distance r from the origin for small and large values of x and y can determine the divergent or convergent properties of the phase trajectories. Several other conditions for the existence of limit cycles have been determined.⁽³⁾ Some of these conditions are the following:

(a) No limit cycle exists in any region within which

$$\frac{\delta P}{\delta x} + \frac{\delta Q}{\delta y}$$

does not change sign.

(b) Within any limit cycle the number of nodes, foci, and centers must exceed the number of saddle points by one.

(c) If a trajectory stays inside a finite region and does not approach a singular point, then the trajectory must be a limit cycle or approach a limit cycle asymptotically.

Knowing the trajectories, the singular points, and the limit cycles of a system, the behavior of the system is completely determined when the initial conditions are specified. The determination of the trajectories and the singular points is a straightforward

procedure; the determination of the limit cycles, however, is more difficult. The complete phase portrait can then be used to deter-

mine the nature of system stability, and the response of the system (if stable) is readily ascertained.

10-4 LIMITATIONS, COMPENSATION, AND OTHER METHODS

As discussed previously, the describing-function procedure is primarily effective in determining the existence of limit cycles and predicting the amplitudes and frequencies associated with stable limit cycles. To a lesser degree the describing-function procedure can be used to estimate qualitatively the degree of stability of stable nonlinear systems. In addition, the describing-function method can be used to determine the frequency response of nonlinear systems therefore, it is helpful in explaining anomalous experimental results.

The phase-plane procedure is useful in determining the *exact* nature of the stability of nonlinear systems in situations where the describing function method is inapplicable. In addition, the *time response* of a nonlinear system can be determined expeditiously through the use of the phase-plane so that a more quantitative estimate of the degree of stability of a system can be obtained.

Unfortunately, neither the phase plane nor the describing function can be used to determine the response of a nonlinear system to inputs other than simple steps or sinusoids. Since these elementary inputs rarely occur in practice, the utility of the two methods is severely restricted.

When the input to a system is arbitrarily defined, it is necessary to use either numerical computation^(54,55,56) or, more conveniently, analog or digital computers. The analog computer is an especially powerful aid in the study of nonlinear systems.

Some specific remarks are in order regarding the stabilization and compensation of nonlinear systems. If the describing-function method is applicable, stabilization can be accomplished by reshaping the response $G(j\omega)$

of the linear element with conventional linear functions to eliminate intersections between the describing function and $G(j\omega)$. A nonlinear compensation function may be added to reshape the original describing function. If the added nonlinearity is separated from the original one by a low-pass filter, the describing functions of the two nonlinearities can be multiplied directly to obtain the composite describing function of the nonlinearly compensated system. If the compensating nonlinearity immediately precedes or follows the original nonlinearity with no separation by filtering action, a new describing function must be determined by combining the input-output characteristics of the two nonlinearities. In the latter case, the effect of the added nonlinearity on the original describing-function locus is much more difficult to visualize.

A great deal of effort has been devoted to the study of "optimum" nonlinear systems. The basic assumption in these studies^(5,11,23,24,25,27,30,34,41,44,46,48) is that a system having a transient response (to a *step input*) that settles in a minimum period of time and has a minimum overshoot is an "optimum" system. The limitation of such "optimization" methods is due primarily to the fact that a nonlinear system will behave differently for different inputs. As a result, a system that has been "optimized" for a given step input may behave poorly for other step inputs of different magnitude, and it is likely that it will not behave in an optimum manner in response to other types of inputs.

In conclusion, it should be said that the problems of stabilization, compensation, and optimization of nonlinear systems have, as yet, not been adequately treated.

NONLINEAR SYSTEMS

BIBLIOGRAPHY

- 1 H. T. Marcy, M. Yachter, and J. Zauderer, "Instrument Inaccuracies in Feedback Control Systems with Particular Reference to Backlash", *Trans. AIEE*, Vol. 68, Part I, pp. #778-788, 1949.
- 2 H. K. Weiss, "Analysis of Relay Servomechanisms", *J. of Aero. Sci.*, Vol. 13, No. 7, pp. #364-376, July, 1946.
- 3 N. Minorsky, *Introduction to Nonlinear Mechanics*, J. W. Edwards Bros., Ann Arbor, Mich., 1947.
- 4 A. A. Andronow and C. E. Chaikin, *Theory of Oscillations*, Princeton University Press, Princeton, N. J., 1949.
- 5 A. M. Hopkin, "A Phase-Plane Approach to the Compensation of Saturating Servomechanisms", *Trans. AIEE*, Vol. 70, Part I, pp. #631-639, 1951.
- 6 D. A. Kahn, "An Analysis of Relay Servomechanisms", *Trans. AIEE*, Vol. 68, Part II, pp. #1079-1088, 1949.
- 7 R. J. Kochenburger, "A Frequency-Response Method for Analyzing and Synthesizing Contactor Servomechanisms", *Trans. AIEE*, Vol. 69, Part I, pp. #270-284, 1950.
- 8 A. Tustin, "The Effects of Backlash and of Speed-Dependent Friction on the Stability of Closed-Cycle Control Systems", *J. Inst. Elec. Engrs.* (London), Vol. 94, Part IIA, No. 1, pp. #143-151, May, 1947.
- 9 G. D. McCann, F. C. Lindvall, and C. H. Wilts, "The Effect of Coulomb Friction on the Performance of Servomechanisms", *Trans. AIEE*, Vol. 67, Part I, pp. #540-546, 1948.
- 10 C. Leondes and M. Rubinoff, "DINA, A Digital Analyzer for Laplace, Poisson, Diffusion, and Wave Equations", *Trans. AIEE*, Vol. 71, pp. #303-309, 1952.
- 11 J. B. Lewis, "The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism", *Trans. AIEE*, Vol. 71, Part II, pp. #449-453, 1952.
- 12 N. Kryloff and No. Bogoliuboff, *Introduction to Nonlinear Mechanics*, translated by S. Lefschetz, Princeton University Press, Princeton, N. J., 1943.
- 13 B. V. Bulgakov, "Periodic Processes in Free Pseudo-Linear Oscillatory Systems", *J. Franklin Inst.*, Vol. 235, pp. #591-616, June, 1943.
- 14 J. C. Lozier, "Carrier-Controlled Relay Servos", *Electrical Engineering*, Vol. 69, No. 12, pp. #1052-1056, December, 1950.
- 15 E. C. Johnson, "Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements", *Trans. AIEE*, Vol. 71, Part II, pp. #169-181, 1952.
- 16 E. S. Sherrard, "Stabilization of a Servomechanism Subject to Large Amplitude Oscillation", *Trans. AIEE*, Vol. 71, Part II, pp. #312-324, 1952.
- 17 E. Levinson, "Some Saturation Phenomena in Servomechanisms with Emphasis on the Tachometer Stabilized System", *Trans. AIEE*, Vol. 72, Part II, pp. #1-9, 1953.
- 18 V. B. Haas, Jr., "Coulomb Friction in Feedback Control Systems", *Trans. AIEE*, Vol. 72, Part II, pp. #119-126, 1953.

- 19 R. J. Kochenburger, "Limiting in Feed-back Control Systems", *Trans. AIEE*, Vol. 72, Part II, pp. #180-194, 1953.
- 20 W. E. Scott, "An Introduction to the Analysis of Nonlinear Closed-Cycle Control Systems", in *Automatic and Manual Control*, edited by A. Tustin, pp. #249-261, Academic Press, Inc., New York, N. Y., 1952.
- 21 R. L. Cosgriff, "Open-Loop Frequency-Response Method for Nonlinear Servomechanisms", *Trans. AIEE*, Vol. 72, Part II, pp. #222-225, 1953.
- 22 H. D. Greif, "Describing-Function Method of Servomechanism Analysis Applied to Most Commonly Encountered Nonlinearities", *Trans. AIEE*, Vol. 72, Part II, pp. #243-248, 1953.
- 23 R. S. Neiswander and R. H. MacNeal, "Optimization of Nonlinear Control Systems by Means of Nonlinear Feedbacks", *Trans. AIEE*, Vol. 72, Part II, pp. #262-272, 1953.
- 24 D. McDonald, "Nonlinear Techniques for Improving Servo Performance", in *Proc. Natl. Electronics Conf.*, Vol. 6, pp. #400-421, Chicago, Ill., 1950.
- 25 D. McDonald, "Multiple Mode Operation of Servomechanisms", *Rev. Sci. Instr.*, Vol. 23, pp. #22-30, January, 1952.
- 26 S. Lefschetz, *Contributions to the Theory of Nonlinear Oscillations*, Princeton University Press, Princeton, N. J., 1950.
- 27 L. F. Kazda, "Errors in Relay Servo Systems", *Trans. AIEE*, Vol. 72, Part II, pp. #323-328, 1953.
- 28 T. A. Rogers and W. C. Hurty, "Relay Servomechanisms: The Shunt-Motor Servo with Inertia Load", *Trans. ASME*, Vol. 72, pp. #1163-1172, November, 1950.
- 29 A. M. Uttley and P. H. Hammond, "The Stabilization of On-Off Controlled Servomechanisms", in *Automatic and Manual Control*, edited by A. Tustin, pp. #285-307, Academic Press, New York, N. Y., 1952.
- 30 T. M. Stout, "Effects of Friction in an Optimum Relay Servomechanism", *Trans. AIEE*, Vol. 72, Part II, pp. #329-336, 1953.
- 31 N. B. Nichols, "Backlash in a Velocity Lag Servomechanism", *Trans. AIEE*, Vol. 72, Part II, pp. #462-467, 1953.
- 32 J. R. Burnett and P. E. Kendall, "Linear Compensation of Saturating Servomechanisms", *Trans. AIEE*, Vol. 73, Part II, pp. #6-10, 1954.
- 33 C. K. Chow, "Contactor Servomechanisms Employing Sampled Data", *Trans. AIEE*, Vol. 73, Part II, pp. #51-64, 1954.
- 34 I. Bogner and L. F. Kazda, "An Investigation of the Switching Criteria for Higher Order Contactor Servomechanisms", *Trans. AIEE*, Vol. 73, Part II, pp. #118-127, 1954.
- 35 L. T. Prince, Jr., "A Generalized Method for Determining the Closed-Loop Frequency Response of Nonlinear Systems", *Trans. AIEE*, Vol. 73, Part II, pp. #217-224, 1954.
- 36 D. K. Gehmlich and M. E. Van Valkenberg, "Measurement of Some Nonlinearities in Servomechanisms", *Trans. AIEE*, Vol. 73, Part II, pp. #232-235, 1954.
- 37 J. G. L. Michel and A. Porter, "The Effect of Friction on the Behavior of Servomechanisms at Creep Speeds", *Proc. Inst. Elec. Engrs. (London)*, Vol. 98, Part II, pp. #297-311, 1951.

NONLINEAR SYSTEMS

- 38 H. Lauer, "Operating Modes of a Servomechanism with Nonlinear Friction", *J. Franklin Inst.*, Vol. 255, pp. #497-511, 1953.
- 39 Y. H. Ku, "Nonlinear Analysis of Electro-Mechanical Problems", *J. Franklin Inst.*, Vol. 255, pp. #9-31, 1953.
- 40 Y. H. Ku, "A Method for Solving Third and Higher Order Nonlinear Differential Equations", *J. Franklin Inst.*, Vol. 256, pp. #229-244, 1953.
- 41 R. E. Kalman, "Phase-Plane Analysis of Automatic Control Systems with Nonlinear Gain Elements", *Trans. AIEE*, Vol. 73, Part II, pp. #383-390, 1954.
- 42 L. M. Vallese, "Analysis of Backlash in Feedback Control Systems with One Degree of Freedom", *Trans. AIEE*, Vol. 74, Part II, pp. #1-4, 1955.
- 43 K. C. Matthews and R. C. Boe, "The Application of Nonlinear Techniques to Servomechanisms", *Proc. Natl. Electronics Conf.*, Vol. 8, pp. #10-21, Chicago, Ill., 1952.
- 44 J. E. Hart, "An Analytical Method for the Design of Relay Servomechanisms", *Trans. AIEE*, Vol. 74, Part II, pp. #83-90, 1955.
- 45 M. V. Matthews, "A Method for Evaluating Nonlinear Servomechanisms", *Trans. AIEE*, Vol. 74, Part II, pp. #114-123, 1955.
- 46 S. S. L. Chang, "Optimum Switching Criteria for Higher Order Contactor Servo with Interrupted Circuits", *Trans. AIEE*, Vol. 74, Part II, pp. #273-276, 1955.
- 47 Y. H. Ku, "Analysis of Nonlinear Systems with More than One Degree of Freedom by Means of Space Trajectories", *J. Franklin Inst.*, Vol. 259, pp. #115-131, February, 1955.
- 48 R. E. Kalman, "Analysis and Design Principles of Second and Higher Order Saturating Servomechanisms", *Trans. AIEE*, Vol. 74, Part II, pp. #294-310, 1955.
- 49 R. C. Booton, Jr., "The Analysis of Nonlinear Control Systems with Random Inputs", *Proc. of the Symposium of Nonlinear Circuits Analysis*, pp. #369-391, Brooklyn Polytechnic Inst., Brooklyn, N. Y., April, 1953.
- 50 K. Chen, "Quasi-Linearization Techniques for Transient Study of Nonlinear Feedback Control Systems", *Trans. AIEE*, Vol. 74, Part II, pp. #354-365, 1955.
- 51 H. Chestnut and R. W. Mayer, *Servomechanisms and Regulating System Design*, Vol. II, pp. #241-364, John Wiley & Sons, Inc., New York, N. Y., 1955.
- 52 J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. #559-663, McGraw-Hill Book Company, Inc., New York, N. Y., 1955.
- 53 I. Flügge-Lotz, *Discontinuous Automatic Control*, Princeton University Press, Princeton, N. J., 1953.
- 54 A. S. Householder, *Principles of Numerical Analysis*, McGraw-Hill Book Company, Inc., New York, N. Y., 1953.
- 55 F. B. Hildebrand, *Introduction to Numerical Analysis*, McGraw-Hill Book Company, Inc., New York, N. Y., 1956.
- 56 W. E. Milne, *Numerical Solutions of Differential Equations*, John Wiley & Sons, Inc., New York, N. Y., 1953.

AMC PAMPHLET

AMCP 706-137

THIS IS A REPRINT WITHOUT CHANGE OF ORDP 20-137, REDESIGNATED AMCP 706-137

ENGINEERING DESIGN HANDBOOK

SERVOMECHANISMS

SECTION 2, MEASUREMENT AND SIGNAL CONVERTERS



HEADQUARTERS
UNITED STATES ARMY MATERIEL COMMAND
WASHINGTON, D.C. 20315

30 April 1965

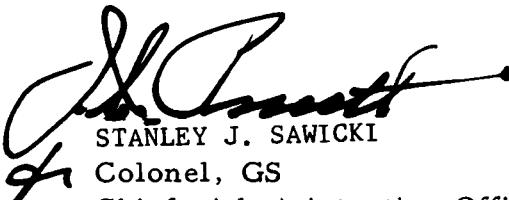
AMCP 706-137, Servomechanisms, Section 2, Measurement and Signal Converters, forming part of the Army Materiel Command Engineering Design Handbook Series, is published for the information and guidance of all concerned.

(AMCRD)

FOR THE COMMANDER:

SELWYN D. SMITH, JR.
Major General, USA
Chief of Staff

OFFICIAL:



STANLEY J. SAWICKI
Colonel, GS
Chief, Administrative Office

DISTRIBUTION: Special

PREFACE

The Engineering Design Handbook Series of the Army Materiel Command is a coordinated series of handbooks containing basic information and fundamental data useful in the design and development of Army materiel and systems. The handbooks are authoritative reference books of practical information and quantitative facts helpful in the design and development of Army materiel so that it will meet the tactical and the technical needs of the Armed Forces. The present handbook is one of a series on Servomechanisms.

Section 2 of the handbook contains Chapters 11 and 12, which describe those servomechanism components used as sensing elements and signal converters. Chapter 11 covers various sensing elements such as potentiometers, rotary transformers, linear variable differential transformers, tachometer generators, gyroscopes and analog-to-digital converters. Chapter 12 covers three types of signal converters: modulators, demodulators and digital-to-analog converters.

For information on other servomechanism components and on feedback control theory and system design, see one of the following applicable sections of this handbook:

- AMCP 706-136 Section 1 Theory (Chapters 1-10)
- AMCP 706-138 Section 3 Amplification (Chapter 13)
- AMCP 706-139 Section 4 Power Elements and System Design (Chapters 14-20)

An index for the material in all four sections is placed at the end of Section 4.

Elements of the U. S. Army Materiel Command having need for handbooks may submit requisitions or official requests directly to Publications and Reproduction Agency, Letterkenny Army Depot, Chambersburg, Pennsylvania 17201. Contractors should submit such requisitions or requests to their contracting officers.

Comments and suggestions on this handbook are welcome and should be addressed to Army Research Office-Durham, Box CM, Duke Station, Durham, North Carolina 27706.

CHAPTER 11

SENSING ELEMENTS

11-1 INTRODUCTION*

This chapter deals with those devices that measure or sense the input or output of a servomechanism and express it in electrical form. The more common types of sensing elements used in ordnance applications convert motion, either translational or rotational, into a corresponding electrical representation. They are part of a class of devices called *transducers*, the function of a transducer being to receive information from some source and to express that information in a different form for use by another operating component.

The variety of sensing elements that have

been developed is very large. No attempt is made here, however, to list or discuss all of them. Rather, the selection is restricted to those that find wide use in ordnance applications.

In the remainder of this chapter, emphasis is placed on discussing the characteristics and limitations of the various sensing elements considered. The aim is to provide the servo designer with sufficient information to select a suitable type of sensing element for a given application and to match it properly to the rest of the system utilizing the servo.

11-2 POTENTIOMETERS*

11-2.1 DESCRIPTION AND BASIC THEORY

11-2.2 Definition

A potentiometer is a resistor with a sliding contact. The electrical resistance between the slider and either end point of the total resistance element is a predetermined function of the distance of the slider from the end point so that the potentiometer converts slider position into electrical resistance.

Figure 11-1 shows the electrical representation of a potentiometer. The arrow indicates the movable slider; A and B are the end points of the total resistance element across which the excitation voltage is applied; C and D are the end points of that portion of the total resistance element across which the

variable output voltage is taken. In an actual potentiometer, one terminal suffices for end points B and D combined.

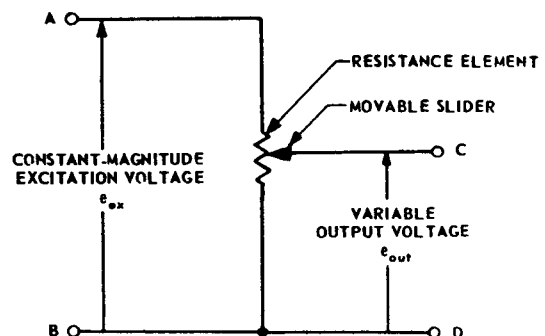


Fig. 11-1 Electrical representation of a potentiometer.

*By A.K. Susskind

11-2.3 Types of Potentiometers

There are two basic types of potentiometers:⁽¹⁾ *rotary* and *translatory*. The majority of potentiometers produced today are intended for applications where the slider motion is rotary. The basic rotary type can be broken down into two subtypes: the *single-turn* potentiometer and the *multiturn* potentiometer.

Single-turn potentiometers are designed for slider-travel limits of a full revolution or less (e.g., 300°). Multiturn potentiometers are designed for slider-travel limits of several revolutions (e.g., some multiturn potentiometers are now being made with as many as 60 revolutions between stops).

11-2.4 Principle of Operation

The basic performance equation for a single-turn or multiturn potentiometer can be written⁽¹⁾

$$R(\theta) = \int_0^\theta \rho(\theta) d\theta \quad (11-1)$$

where

θ = slider angle (i.e., the angular displacement of the slider with respect to the zero end of the potentiometer)

$R(\theta)$ = resistance between the slider and the zero end of the potentiometer (i.e., the resistance between points C and D in Fig. 11-2)

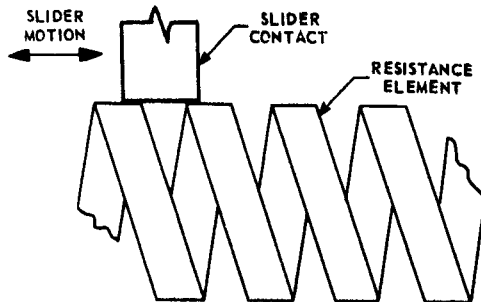


Fig. 11-2 Wire-wound element and slider.

$\rho(\theta)$ = change in resistance per unit angular displacement as a function of slider angle θ (in units consistent with those chosen for θ)

The total winding resistance R_w can be written

$$R_w = \int_0^{\theta_{max}} \rho(\theta) d\theta \quad (11-2)$$

where θ_{max} is the slider angle that represents the angular displacement of the slider from the zero end of the potentiometer to the other end.

If $\rho(\theta)$ is a constant, k_1 , then

$$R(\theta) = k_1\theta \quad (11-3)$$

and

$$R_w = k_1\theta_{max} \quad (11-4)$$

When $\rho(\theta)$ is constant, the potentiometer is known as a *linear* potentiometer. If $\rho(\theta)$ is not constant, the potentiometer is said to be *nonlinear*. The form of Eqs. (11-1) and (11-2) indicates that the resistivity function $\rho(\theta)$ of a potentiometer must be the derivative of the desired resistance function; i.e.,

$$\rho(\theta) = \frac{dR(\theta)}{d\theta} \quad (11-5)$$

When an excitation voltage e_{ex} of fixed amplitude is applied across the potentiometer, the output voltage e_{out} measured between the zero end of the potentiometer and the slider is a function of slider position; i.e.,

$$e_{out} = e_{ex} \frac{R(\theta)}{R_w} \quad (11-6)$$

The relationships for a translatory potentiometer that correspond to Eqs. (11-1) through (11-6) can be written directly as follows:

$$R(x) = \int_0^x \rho(x) dx \quad (11-7)$$

$$R_w = \int_0^{x_{max}} \rho(x) dx \quad (11-8)$$

$$R(x) = k_2 x \quad (11-9)$$

$$R_w = k_2 x_{max} \quad \left. \vphantom{R(x)} \right\} \text{for } \rho(x) = \text{constant} = k_2 \quad (11-10)$$

$$\rho(x) = \frac{dR(x)}{dx} \quad (11-11)$$

$$e_{out} = e_{ex} \frac{R(x)}{R_w} \quad (11-12)$$

where

x = slider displacement (i.e., the linear distance of the slider from the zero end of the potentiometer)

$R(x)$ = resistance between the slider and the zero end of the potentiometer

$\rho(x)$ = change in resistance per unit linear displacement as a function of slider displacement x (in units consistent with those chosen for x)

x_{max} = distance represented by slider displacement from the zero end of the potentiometer to the other end

K_2 = constant value of $\rho(x)$

11-2.5 Use

An important application of potentiometers in servos is as an electromechanical transducer for converting a mechanical displacement into a corresponding voltage signal. A potentiometer is one of the simplest means of accomplishing this function. If a direct or alternating excitation voltage of fixed amplitude is connected across the potentiometer, the output voltage (measured between the slider and the zero-end terminal) is a function of slider position and is given by either Eq. (11-6) or whichever is applicable. It follows that only when a linear potentiometer is used will the output be a linear function of slider position. Since a linear sensing element is required in most servomechanisms, linear potentiometers are nearly always used and the required tolerance in $R(\theta)$ is so close that only precision potentiometers are employed

Nonlinear potentiometers are used primarily in computing circuits. For example, by making

$$\rho(\theta) = \cos \theta \quad (11-13)$$

the resistance $R(\theta)$ [see Eq. (11-1)] becomes

$$R(\theta) = \int_0^\theta \cos \theta \, d\theta = \sin \theta \quad (11-14)$$

Hence, this particular nonlinear potentiometer can be used to express electrically the sine of the slider position.

11-2.6 Construction Features

A potentiometer consists of three main parts: the resistance element; the slider; and the housing. The housing holds the other two parts in proper relationship to each other and also serves as a mount for the complete unit. In rotary-motion units, a shaft for mechanical coupling to the slider is brought out through the housing; in linear-motion units, a rod is brought out.

Resistance elements of four types are used in precision potentiometers for servomechanism applications: wire wound; slide wire; film; and conductive plastic. Film elements (made of carbon or metal) have not yet been developed as fully as wired-wound and slide-wire elements. At the present time, most commercially available precision potentiometers contain wire-wound resistance elements.

Wire-wound elements are made by wrapping an insulated resistance wire helically around either an insulating card or an insulated metallic rod that, for rotary units, is then bent into a circular shape. The wire insulation is removed along the path that the slider contacts. When moving, the slider successively contacts each turn of the winding as shown in Fig. 11-2. In the position shown, the slider is in contact with two turns of the resistance element. As the slider moves to the right, it leaves the first turn and makes contact with only the second until further motion brings it into contact with both the second and third. The resistance between the slider and one end point of the resistance element therefore varies in discrete steps, as does the output voltage when a fixed voltage is applied across the entire element. The size of the voltage steps is of the order of e_{ex}/n , where n is the number of turns of the resistance wire. The exact size of the steps varies with slider position and structural details.

Because of these voltage steps, the use of a wire-wound potentiometer as an output sensing element in a servomechanism leads to small oscillations around the particular step that most nearly corresponds to the servomechanism command. The more turns the element has, the smaller the amplitude of the oscillations. Hence, in selecting a potentiometer, care must be taken to assure that the size of the resistance steps is small enough to result in oscillations of negligible amplitude, or compensation can be added to the servomechanism in the form of coulomb friction at the output shaft. Gottling⁽¹⁸⁾ has shown that for a simple positional servomechanism consisting of an amplifier and motor, with voltage feedback from the output shaft obtained by means of a wire-wound potentiometer, the output will settle on the wire step where the voltage is nearest to the input if the coulomb friction, q , is in accordance with the following criterion

$$q \geq \frac{1}{2} \times \frac{e_{cs}}{n} \times K_a K_m \quad (11-15)$$

where

K_a = amplifier gain

K_m = stalled motor torque per unit control-phase voltage

For a given potentiometer diameter, the smaller the wire diameter, the larger the total resistance and the greater the number of turns. The greater the number of turns, the smaller the individual voltage steps and the smaller the travel of the slider arm over which the output voltage is constant. Hence, the number of wire turns also determines the accuracy to which a desired voltage can be achieved.

At the present time, the smallest practical wire diameter used is 0.9 mil. In typical commercial units presently available, the maximum case diameter is a few inches and the maximum number of wire turns per shaft revolution is approximately 5000, so that the upper limit in resolution, obtained in the highest-resistance units, is approximately one part in 10,000 per revolution. Where a circuit design calls for a low impedance level

and the inherently poor resolution of a low-resistance potentiometer cannot be tolerated, one can use a high-resistance, good-resolution potentiometer and shunt it across the input terminals with a fixed low resistance. This arrangement provides a low impedance level at the input terminals, without sacrificing the desired resolution.

The stepped nature of the resistance function is the most serious disadvantage of wire-wound elements. It is overcome in the carbon, metal, and conductive-plastic film elements, which represent a nearly smooth, unbroken surface to the slider and exhibit voltage steps smaller than those of a high-resistance, wire-wound element by a factor of several tens. At present, however, only a small variety of film units are commercially available and these are not yet as stable under extreme environmental conditions as are wire-wound elements. (Another form of a smooth resistance element is the composition type. However, its resistivity function cannot be sufficiently well controlled in manufacture to permit its use as a sensing element for servomechanisms.)

While very good resolution can be obtained with potentiometers of the slide-wire type, these units have relatively low total winding resistance. This can result in excessive loading of the electrical source. The total resistance of the element is limited by the size of the potentiometer. Typical single-turn units have a total resistance of a few hundred ohms; typical multiturn units have a total resistance of a few thousand ohms.

11-2.7 LINEAR POTENTIOMETERS

11-2.8 Types of Linearity

The characteristic of greatest interest in a linear potentiometer is the degree of linearity; i.e., the closeness with which the output voltage or resistance is a linear function of slider position. The inherent stepped nature of the output of a wire-wound potentiometer yields a linearity deviation that is one half the size of the maximum step. Manufacturing tolerances add further deviations from perfect linearity.

SENSING ELEMENTS

Figure 11-3 shows a typical plot of percent output voltage as a function of percent slider position, with percent deviation from linearity $(D)E_{out}/E_{ox}$ exaggerated. The maximum percent deviation of the actual curve from the ideal curve is called *terminal linearity*. Terminal linearities as low as 0.01 percent have been achieved, but linearities of 0.1 to 0.5 percent are more common in most commercial units.

A different measure of potentiometer performance commonly used is *independent linearity*, which is defined by the Radio, Electronic and Television Manufacturers' Association (RETMA) as "the deviation when the slope and position of a straight reference line are chosen to make the maximum deviations a minimum over the actual effective travel or any specified portion thereof." The straight reference line used does not coincide with the line corresponding to ideal performance. The independent linearity figure is smaller than the terminal linearity figure.

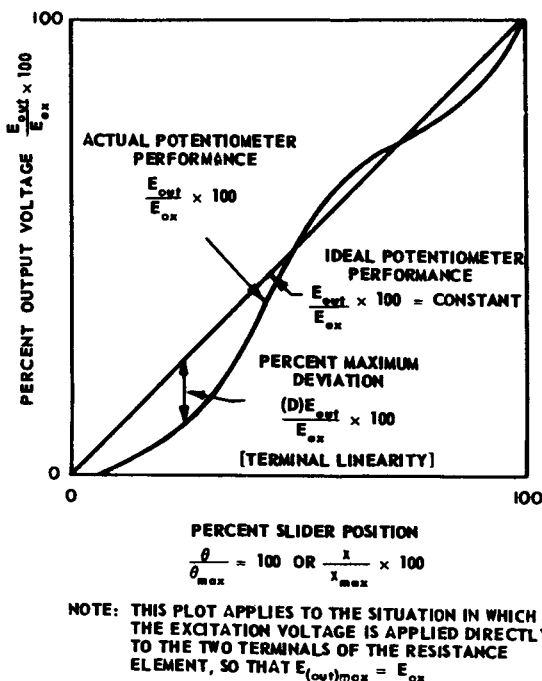


Fig. 11-3 Potentiometer linearity performance compared with ideal performance.

The best straight line for the actual characteristics previously shown is drawn in Fig. 11-4, which also shows that the maximum deviation now occurs at a different slider position and is smaller than that in Fig. 11-3. Note that the best straight line does not pass through the origin nor does it intersect the actual performance curve at the 100-percent point. The best straight line can be made to coincide with the required straight line by trimming techniques, so that the actual performance deviation from an ideal straight line will differ by only the independent linearity figure. Trimming is accomplished by referring the slider voltage not to the voltage applied across the resistance element but to a value that corresponds to DE in Fig. 11-4.

A typical trimming circuit is shown in Fig. 11-5. The upper and lower trimmer potentiometers are center-tapped and the voltage between the two center taps is taken as the reference. The sliders on the two trimmer potentiometers are adjusted so that when the slider of the precision potentiometer is at each end, the percentage output voltage differs from the best straight line by m and n , respectively, of Fig. 11-4.

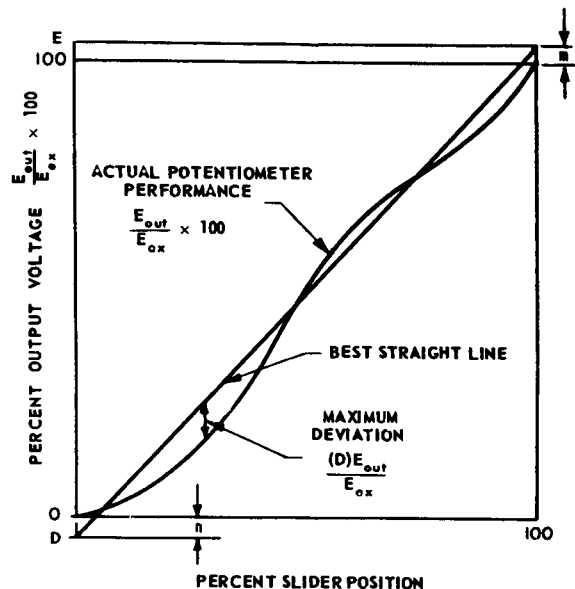


Fig. 11-4 Potentiometer linearity performance compared with the best straight line.

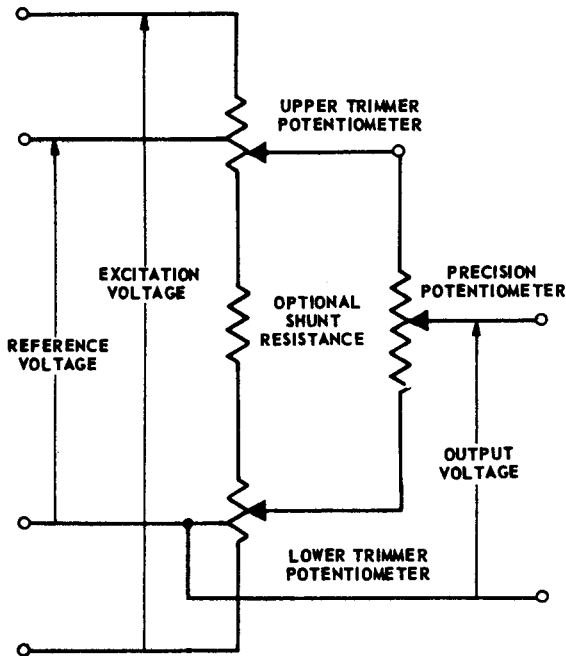


Fig. 11-5 Potentiometer with a typical trimming circuit.

Another definition of linearity that is sometimes used is *zero-based linearity*. This is defined by the RETMA as “the deviation from a straight reference line through zero applied voltage as the starting end point, with its slope chosen to make the maximum deviation minimum over the theoretical effective travel or any portion thereof.” The addition of a trimming circuit at the upper end of the potentiometer can cause the actual performance to differ from a straight line by no more than the zero-based linearity.

11-2.9 Effect of Load Impedance

In a circuit application, the linearity of the output voltage is a function not only of the potentiometer linearity itself, but also of the load resistance that is connected to the slider. The larger the load resistance with respect to the total potentiometer resistance, the smaller the deviation of the output voltage from the potentiometer linearity curve. The output voltage in the presence of load-

ing, assuming an ideal linear potentiometer, is⁽²⁾

$$E_{out} = E_{ex} \frac{(R_l/R_w) [R(\theta)/R_w]}{R_l/R_w + R(\theta)/R_w [1 - R(\theta)/R_w]} \quad (11-16)$$

where

R_l = load resistance

$R(\theta)$ = resistance between the slider and the zero end of the potentiometer

R_w = total winding resistance of the potentiometer

A plot of the ratio of the output voltage to the excitation voltage is given in Fig. 11-6 as a function of the ratio $R(\theta)/R_w$ for various values of the ratio R_l/R_w .

There are several methods of reducing the error due to loading. One means, applicable to servomechanisms deriving the input signal from a potentiometer, is to use identical potentiometers for both the input and the output and to load both by the same amount.

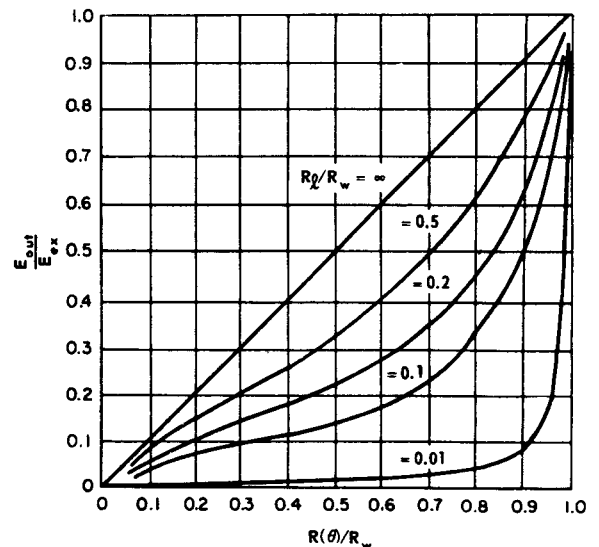


Fig. 11-6 Effect of loading on potentiometer output.

Adapted by permission from *Electronic Instruments*, MIT Radiation Laboratory Series, Vol. 21; edited by I. A. Greenwood, Jr., J. V. Holdman, Jr., and D. MacRae, Jr.; Copyright 1948; McGraw-Hill Book Company, Inc.

SENSING ELEMENTS

A second method consists of adding a series resistor R_s as shown in Fig. 11-7. Nettleton and Dole⁽³⁾ have shown that for $R_t \gg R_w$ (by at least a factor of 10) the output voltage of the uncompensated circuit has the following maximum error, expressed as a percent of the maximum output voltage:

$$\text{percent maximum error} = \frac{100R_w}{11R_t + R_w} \quad (11-17)$$

When R_s with a value of $0.28R_w$ is added, however, the maximum error expressed as a percent of the maximum output voltage is

$$\text{percent maximum error} = \frac{17.86R_w}{9R_t + 2R_w} \quad (11-18)$$

A reduction in the loading error in the ratio of about four- or five-to-one can thus be accomplished. However, in order to achieve a desired output voltage for $R(\theta) = R_w$, the supply voltage for the compensated case must be made higher than in the uncompensated case. For $R_s = 0.28R_w$, the supply voltage must be 28 percent higher than the desired output voltage for $R(\theta) = R_w$.

A third method of load compensation consists of using a tapped potentiometer and a parallel resistor R_p as shown in Fig. 11-8. Gilbert⁽⁴⁾ has shown that when R_p is $0.31R_t$ and the tap is located so that tap resistance

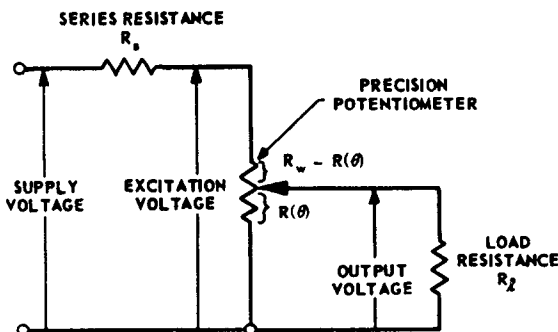


Fig. 11-7 Load compensation by the addition of a series resistor.

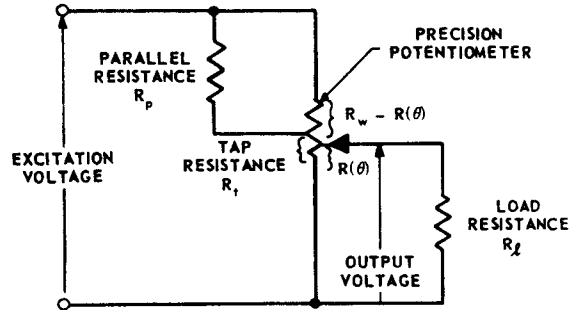


Fig. 11-8 Load compensation by the addition of a potentiometer tap and a parallel resistor.

R_t is $0.74R_w$, the maximum deviation from linearity between the end points of the potentiometer is approximately

$$(D)E_{out} = 0.02 \frac{R_w}{R_t} E_{ex} \quad (11-19)$$

for $R_t \gg R_w$ (by at least a factor of 10). The maximum deviation without tap compensation is over seven times greater.

Table 11-1⁽⁴⁾ shows the results of combining series and tap compensation. The deviation of the output voltage from linearity, $(D)E_{out}/E_{(out) max}$, is the deviation voltage divided by the voltage across the precision potentiometer when the slider is at its maximum displacement. This table is applicable for R_t greater than R_w by a factor of at least 10.

Table 11-2 gives this deviation from linearity between end points when an optimum tap point and a parallel resistor are used.

11-2.10 NONLINEAR POTENTIOMETERS

Nonlinear potentiometers of two types are manufactured. In the first type, the resistivity $\rho(\theta)$ varies almost smoothly, so that $R(\theta)$ is a nearly smooth function. This type is restricted to functions where $\rho(\theta)$ need not vary by a factor of more than 5 over the length of the element. The other type of nonlinear potentiometer approximates the required function $R(\theta)$ by a series of straight lines, $\rho(\theta)$ therefore varying in steps. Larger variations in $\rho(\theta)$ can be achieved with this type than with the smooth type of unit.

TABLE 11-1 LOAD COMPENSATION BY THE ADDITION OF BOTH SERIES AND PARALLEL RESISTORS†

| $\frac{R_s}{R_w}$ | $\frac{R_p}{R_t}$ | $\frac{R_t}{R_w}$ | $\frac{(D)E_{out}}{E_{(out)max}}$ |
|-------------------|-------------------|-------------------|-----------------------------------|
| 0.10 | 0.338 | 0.77 | $\frac{0.017}{(R_t/R_w)}$ |
| 0.25 | 0.255 | 0.86 | $\frac{0.022}{(R_t/R_w)}$ |

$$\dagger \quad \frac{R_t}{R_w} > 10$$

Adapted by permission from *Control Engineering*, Volume 2, No. 2, February, 1955, from article entitled 'Here's a Shortcut in Compensating Pot Loading Errors', by J. Gilbert.

TABLE 11-2 LOAD COMPENSATION BY THE ADDITION OF A PARALLEL RESISTOR

| $\frac{R_t}{R_w}$ | $\frac{R_t}{R_w}$ | $\frac{R_p}{R_t}$ | $\frac{(D)E_{out}}{E_{(out)max}}$ |
|-------------------|-------------------|-------------------|-----------------------------------|
| 1 | 0.73 | 0.281 | 0.025 |
| 2 | 0.73 | 0.304 | 0.011 |
| 3 | 0.73 | 0.315 | 0.0067 |
| 5 | 0.735 | 0.312 | 0.0038 |
| 10 | 0.74 | 0.305 | 0.002 |
| over 10 | 0.74 | 0.311 | $\frac{0.019}{(R_t/R_w)}$ |

Adapted by permission from *Control Engineering*, Volume 2, No. 2, February, 1955, from article entitled 'Here's a Shortcut in Compensating Pot Loading Errors', by J. Gilbert.

The term conformity is used to describe the degree to which the nonlinear potentiometer approximates the desired nonlinear function. *Independent conformity* is the maximum percentage deviation, with respect to the excitation voltage, of the actual electrical output at any point from the best specified function curve drawn through the out-

put versus rotation data. *Terminal conformity* is the maximum percentage deviation, with respect to the excitation voltage, of the actual electrical output at any point from the specified function curve drawn through the end points. By use of trimmer potentiometers similar to those discussed under linear potentiometers, the performance of a nonlinear potentiometer can be made to coincide with its independent conformity.

Nonlinear as well as linear potentiometers are subject to loading errors, and compensation for nonlinear units is accomplished by the same circuits as those discussed under linear potentiometers. For example, Gilbert⁽⁵⁾ shows that where parallel compensation only is used, the tap should be located at 74 percent of total winding resistance R_w , and parallel resistor R_p (connected between the tap and the upper end of the total winding resistance) should be 31 percent of load resistance R_t . However, since nonlinear potentiometers very often must be made up specially for a specific application, the custom design can take into account the effect of a specified load and compensate for it in the element construction.

A tapped linear potentiometer can be used to approximate nonlinear functions by the addition of shunting resistors. This technique is described as follows:

"The first step in the design of a shunting circuit is to draw the required nonlinear function as shown in Fig. 11-9. A series of connected straight lines are next drawn to best represent the function. If the tapped potentiometer has already been constructed, the straight lines must join at angular points corresponding to the tap points. If the tapped potentiometer is to be custom-designed for the particular nonlinear application, the best straight-line approximation can be drawn and the tap points placed at the resulting points of intersection. It is usually necessary to space the taps closely in the region of maximum function curvature.

"An important characteristic of the function of Fig. 11-9 is that the slope does not change its algebraic sign throughout the entire function. The curve commences with a positive slope and continues positive throughout the remainder of the function.

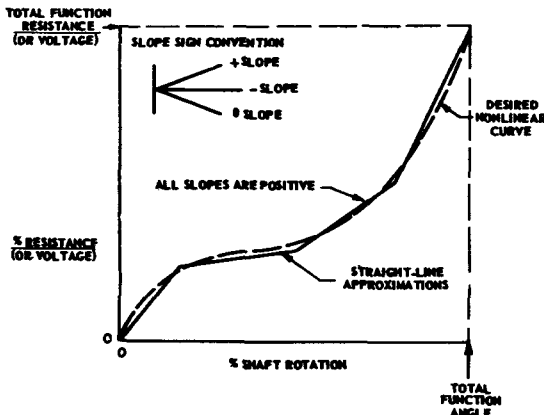


Fig. 11-9 Straight-line approximation of smooth nonlinear function — no change in slope sign.

Reprinted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

"In contrast, the nonlinear function of Fig. 11-10 changes its slope sign twice and thus has three separate regions of unchanging slope sign. Also indicated on Fig. 11-10 are the resistance increments ΔR and shaft angle increments $\Delta\theta$ defined by the straight-line approximation to the curve. These resistance and angle increments are used in the calculation of shunt resistance values.

"The next step is to draw a resistance diagram corresponding to the function diagram of Fig. 11-10. In the graphical representation of Fig. 11-11A, it can be seen that the potentiometer resistance element follows exactly the straight-line approximation of Fig. 11-10. A shunt resistance R_s is connected across each of the three segments of the potentiometer resistance element. End resistors R_e are required at each end of the winding to provide the necessary low-end and high-end voltages. At the points where there is reversal in the algebraic sign of the function, 'pull-in' resistors R_p are used. The resistance diagram can now be redrawn into the more conventional circuit of Fig. 11-11B.

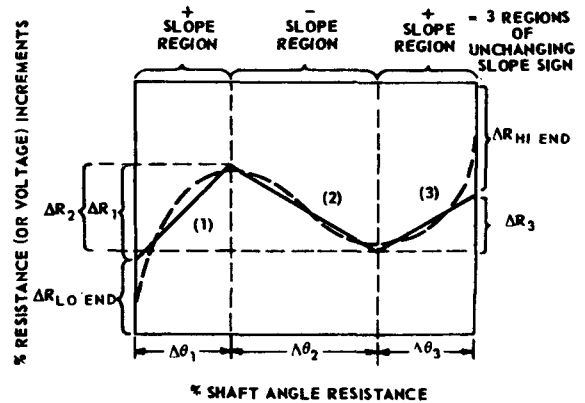


Fig. 11-10 Resistance and shaft angle increments for straight-line approximations to nonlinear function — 3 regions of unchanging slope design.

Reprinted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

"In the calculation of resistors R_s , R_e , and R_p , it is helpful to break the complex network problem into problems of a simpler form. This is done by drawing a separate resistance branch for each region of unchanging slope sign. This results in the three resistance branches of Fig. 11-12. The network division is performed simply by replacing each pull-in resistor R_p with a pair of shunt resistors of value $2R_p$. This form of network division results in equal branch currents i_b flowing in each of the separate resistance branches. Since the same excitation voltage appears across all three branches, the total resistance of each branch, R_b , must be the same.

"This latter condition facilitates the design of a potentiometer shunting network which will have a final resistance, measured between the excitation terminals, of a predetermined value. If the desired resistance of the final shunted potentiometer is called R_d , it can be seen that the branch resistance R_b is simply $R_b = B \times R_d$, where B is the number of resistance branches corresponding to the number of regions of unchanging slope sign. This relation means simply that each resistance branch of Fig. 11-12 must have a total resistance 3 times greater than the desired resistance of the final shunted potentiometer.

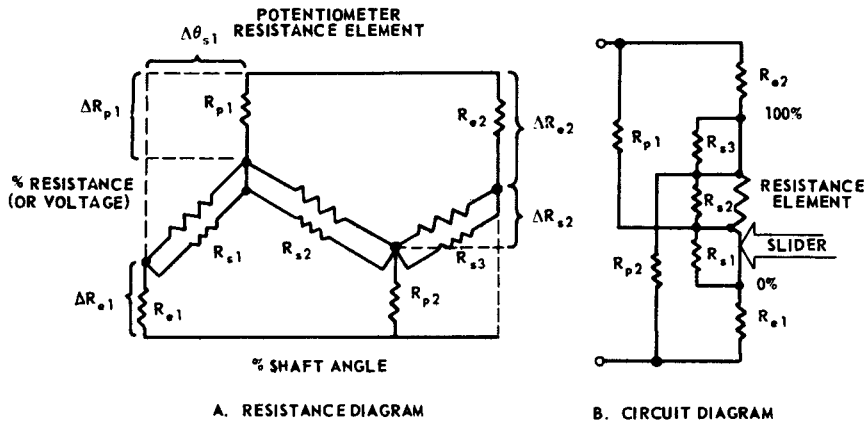


Fig. 11-11 Resistance diagram and resultant circuit diagram for the tapped nonlinear function of Fig. 11-10.

Reprinted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

"To achieve a desired value resistance R_d , it is essential that the winding resistance of the potentiometer to be shunted be greater than a certain minimum value. This results from the fact that shunt resistors can only reduce the net resistance, and the potentiometer winding resistance must be chosen high enough to allow for this reduction effect.

"The minimum allowable winding resistance $R_{w(min)}$ can be determined from the straight-line segment having the greatest slope $(\Delta R/\Delta \theta)_{max}$. The equation relating these variables is

$$R_{w(min)} = \left(\frac{\Delta R}{\Delta \theta} \right)_{max} \times B \times R_d \quad (11-20)$$

It is necessary only to choose the potentiometer having a winding resistance R_w that is greater than this defined minimum value.

"The steps required in the design of a potentiometer shunting network can be summarized as follows:

(a) Plot the desired function $\%R$ (or $\%V$) versus $\%\theta$. These percentages are commonly expressed in decimal form with values ranging from 0 to 1.0.

(b) Approximate the desired function with connected straight lines, joined at selected tap points or at tap points predetermined by the potentiometer construction.

(c) Determine the number of regions of the function having unchanging slope sign. This defines the number of resistance branches B (Fig. 11-12). For functions such as that of Fig. 11-9, B simply equals unity.

(d) Draw a resistance diagram (Fig. 11-11) and define the ΔR and $\Delta \theta$ increments. Label the required shunt resistors R_s , the end resistors R_e , and the pull-in resistors R_p .

(e) Select the desired resistance R_d for the final shunted potentiometer.

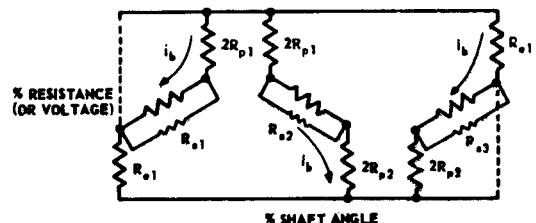


Fig. 11-12 Equivalent resistance diagram showing a resistance branch for each region of unchanging slope sign.

Reprinted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

(f) Determine the value of slope of the straight-line segment having the maximum slope.

(g) Calculate the minimum allowable value of potentiometer winding resistance $R_{ic(min)}$ that is capable of producing the desired function with the specified final resistance R_d [see Eq. (11-19)].

(h) Select a tapped potentiometer having a winding resistance greater than this minimum value, $R_w > R_{ic(min)}$.

(i) Calculate the required end, pull-in, and shunt resistors from the relations

$$R_r = \Delta R_e \times B \times R_d \quad (11-21)$$

$$R_p = \frac{\Delta R_p \times B \times R_d}{2} \quad (11-22)$$

$$\frac{1}{R_s} = \frac{1}{\Delta R_s \times B \times R_d} - \frac{1}{\Delta \theta_s R_w} \quad * \quad (11-23)$$

*Quoted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

Nonlinear functions can also be approximated by loading the slider and adding series resistors to a linear potentiometer. Some typical configurations and the corresponding transfer functions are given in Table 11-3.

Another method of generating nonlinear functions through the use of tapped linear potentiometers is based on clamping the taps at the voltage levels that the desired function has at these points. This is illustrated in Fig. 11-13, which shows the resultant straight-line approximation of the desired function. It is best to derive the clamping voltages from a very low impedance bleeder as shown, for not only does this eliminate the need for many separate bias voltages, but it permits adjustment of the tap voltages when the slider load resistance is connected. If the load resistance is low, the resultant voltage segments between tap points will be concave and can be made to coincide with the desired function more closely by increasing the tap voltages. If the function slope between taps is steep, care must be exercised that the resultant

large voltage drop does not result in currents greater than the potentiometer rating. Thus, the output-voltage scale is determined primarily by the greatest function slope required and the current rating of the potentiometer.

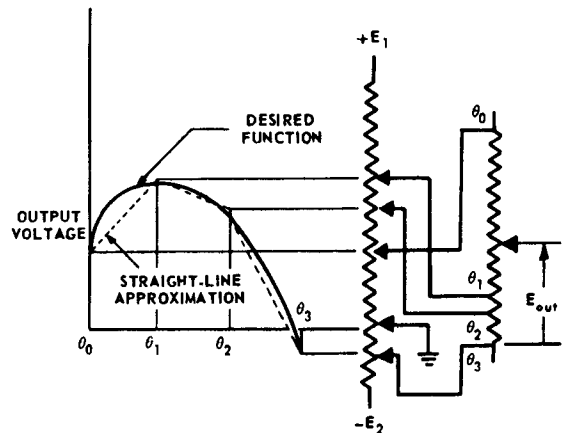


Fig. 11-13 Tapped potentiometer for generating a nonlinear function.

11-2.11 APPLICATION FACTORS

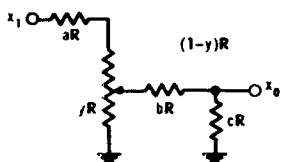
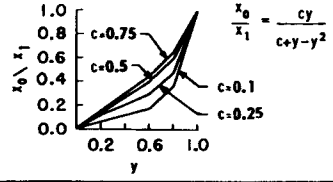
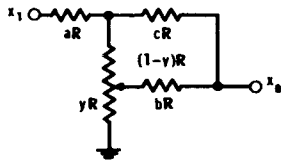
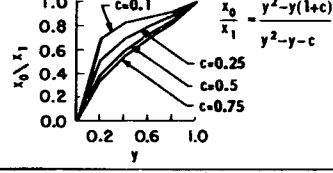
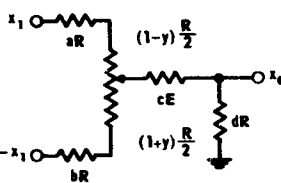
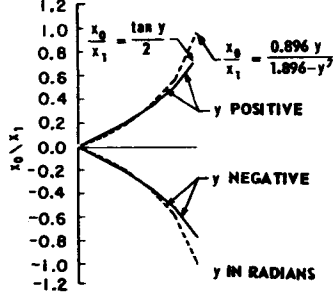
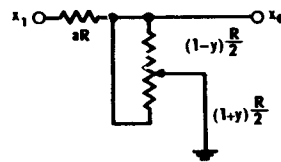
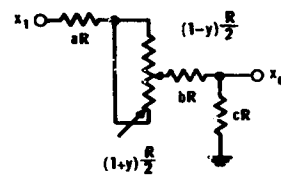
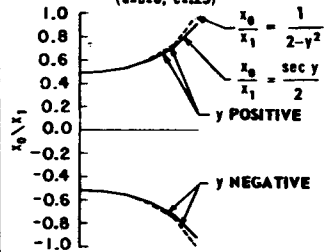
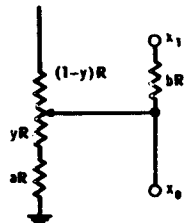
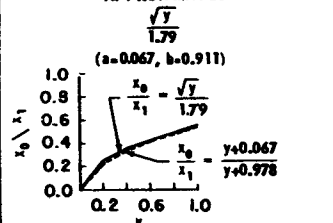
11-2.12 Noise

There are two types of noise voltages, usually not exceeding a few hundred microvolts, that appear at the output of a potentiometer: *active* and *passive*. The former is due to the motion of the slider over the resistance element. The latter is due primarily to the fluctuating contact resistance between the slider and the resistance element and may be reduced by increasing the pressure of the slider against the element. However, this will shorten the life of the potentiometer and increase the driving torque.

Total noise is expressed as an equivalent noise resistance ENR , which is measured as shown in Fig. 11-14. This figure, together

MEASUREMENT AND SIGNAL CONVERTERS

TABLE 11-3 APPROXIMATION OF NONLINEAR FUNCTIONS THROUGH SLIDER LOADING AND SERIES RESISTORS

| CIRCUIT | TRANSFER FUNCTION | TYPICAL FUNCTIONS |
|---|--|--|
|  | $\frac{cy}{(b+c)(a+1)+(a+1)y-y^2}$ |  |
|  | $\frac{y^2-y(1+c)-b}{y^2-y(1-a)-a-(a+1)(b+c)}$ |  |
|  | $\frac{4(b+y-a)d}{[2(a+b)+4(c+d+ab+bd+ad+ac+bc)+1]+2(a-b)y-y^2}$ | <p>APPROXIMATES Tan y (a=b=c=0; d=.224)</p>  |
|  | $\frac{1-y^2}{4a+1-y^2}$ | <p>APPROXIMATES $\frac{1-y^2}{4a}$ a LARGE</p> |
|  | $\frac{4c}{4(a+b+c)+1-y^2}$ | <p>APPROXIMATES Sec y (a=b=0; c=.25)</p>  |
|  | $\frac{y+a}{y+a+b}$ | <p>APPROXIMATES $\frac{\sqrt{y}}{1.79}$ (a=0.067, b=0.911)</p>  |

with the following explanation, is taken from Altieri.⁽⁶⁾

"A source of constant current is placed across one arm of the potentiometer and the slider. The slider and the other arm of the potentiometer are joined to a voltage-measuring circuit having a high input impedance. Owing to the high impedance of the voltage output circuit, the entire source current i_s flows through the left-hand portion of the potentiometer and through the slider.

"If the source current is zero, the noise voltage measured at the circuit output can be attributed entirely to the active noise potential at the slider contact. When the source current is applied, a voltage drop appears across any passive resistance which may exist at the slider point of contact. This passive noise voltage adds directly to the active noise voltage.

"The total voltage output of the circuit is the sum of active and passive residual noise components. Although the two types of noise are of quite different nature (one is a generated voltage and the other is a passive resistance) it is helpful to express both active and passive noise components as ohmic resistances. The total of these resistances is called the *equivalent noise resistance* and is defined as the total residual noise voltage divided by the source current (Fig. 11-14).

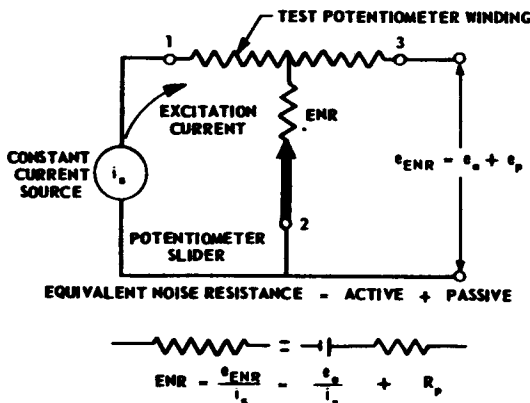


Fig. 11-14 Circuit for measuring equivalent noise resistance, ENR.

By permission from *Instruments*, Volume 26, No. 11, November, 1953, from article entitled 'Residual Potentiometer Noise', by J. R. Altieri.

"The equivalent passive noise resistance appears directly as resistance R_p . The equivalent active noise resistance, however, is the ratio of the active noise voltage (e_a) to the source current (i_s). In standardization of equivalent-noise-resistance measurements, the source current has been arbitrarily set at 1 milliampere flowing from the contact to the winding, and the standard rate of slider rotation has been chosen to be 4 rpm."*

*Quoted by permission from *Instruments* Vol. 26, No. 11, November, 1953, from article entitled "Causes and Measurement of Residual Potentiometer Noise" by J. R. Altieri.

11-2.13 Power Rating

The power rating of a potentiometer states the maximum recommended power that it can dissipate continuously and still satisfy all performance specifications. This power rating is usually specified at a given ambient temperature. Above this temperature, the power rating must be derated. For example, in the instance of potentiometers for which the power rating is given at 40°C, it is customary to apply a linear derating curve with zero dissipation at approximately 85°C.

If a potentiometer is used as a rheostat, the maximum permissible power dissipation must also be derated. Typical derating curves are given in Figs. 11-15 and 11-16. Because of the better heat-conducting characteristics of metal, the derating curve for a metal-base potentiometer is not as severe as for a bakelite-base potentiometer.

The power ratings of potentiometers suitable as sensing elements range from 0.1 to 10 watts. Small units (a fraction of an inch in diameter) have power ratings in the lower end of the range and large units (several inches in diameter) have power ratings in the upper end of the range.

11-2.14 Environmental Effects

While most potentiometers are intended for use at temperatures no greater than 40°C, units are now available that can withstand temperatures as high as 200°C. In selecting a

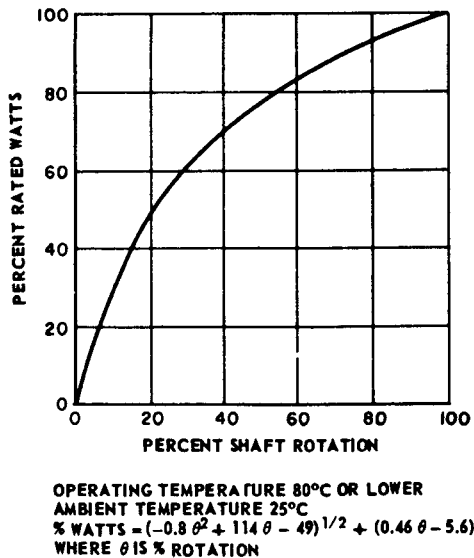


Fig. 11-15 Wattage derating curve for rheostat-connected metal-base potentiometers.

Reprinted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

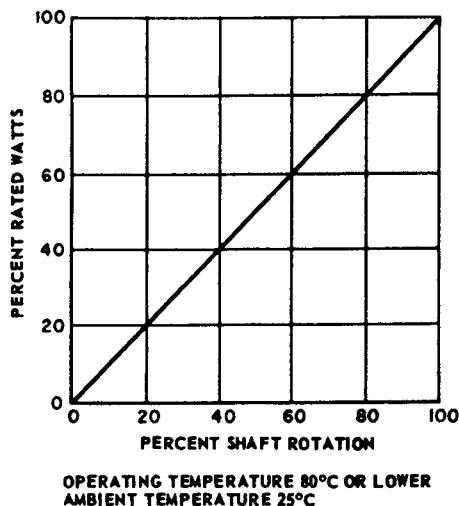


Fig. 11-16 Wattage derating curve for rheostat-connected bakelite-base potentiometers.

Reprinted by permission from the *Potentiometer Handbook*, 1956, Technology Instrument Corp.

potentiometer, care must be taken to select one that can withstand the highest expected temperature, which is a function of the ambient temperature as well as the presence of any nearby heat sources. It must be remembered that temperature affects the total element resistance. From a knowledge of the temperature coefficient of the element and the expected operating temperature (which is a function of the ambient temperature, the presence of heat sources, and the power dissipation in the potentiometer), the resultant element resistance can be computed.

Humidity, fungus, salt spray, and altitude are also factors that must be taken into consideration when selecting a potentiometer. Potentiometer components can be protected against the effects of these factors to various degrees, with hermetically sealed units offering the greatest protection.

Acceleration is another environmental factor that affects the choice of a potentiometer. Units have been designed to withstand acceleration as high as 10 g's for frequencies up to 500 cps.

11-2.15 Life

The life of a potentiometer depends upon the condition of its use. It does not follow that the less the slider rotates the longer the life, for slider motion helps keep the surface between slider and resistance element free of oxide and dirt particles. Well-made potentiometers have been cycled continuously several million times, and many manufacturers now rate the life of their units at one million cycles.

11-2.16 Mechanical Loading

The torque load of the slider on the driving shaft varies with individual designs and ranges from several thousandths of an ounce-inch to several ounce-inches. Most units require about one ounce-inch of starting torque and half that amount for running torque. Moment of inertia figures also vary greatly for different designs and range from less than a thousandth of a gram-square centimeter to several hundred gram-square centimeters.

Most designs result in a slider-assembly inertia of several tens of gram-square centimeters.

11-2.17 Lubrication

In some designs, it has been found desirable to immerse the wire-wound resistance element in a light mineral oil. The flushing

action of the oil reduces electrical noise, disperses wear products which otherwise would tend to lodge between the wire turns, and increases life because of its lubricating action. An example of lubricated potentiometers can be found in the NIKE AJAX Instruction Manual, Vol. VI, Chapter II.

11-3 ROTARY TRANSFORMERS*

11-3.1 GENERAL DESCRIPTION

A rotary transformer is a device in which the coupling between a set of stator coils and a set of rotor poles (not necessarily wound) can be varied by a rotation of either the rotor shaft, the stator assembly, or both. Such a device may be used for the following purposes⁽¹⁹⁾:

(a) To transmit angular information to remote points.

(b) To modulate an electrical signal with mechanical information.

(c) To demodulate an electrical signal and furnish the output information in electrical or mechanical form.

11-3.2 GENERAL CLASSIFICATIONS

Rotary transformers are classified by many different methods according to system, application, construction, basic principles of operation, or manufacturers' trade names.

11-3.3 Use in Positional Systems

In system work, a common use for rotary transformers is the transfer of positional information from one point to another; e.g., in a positional servo where an output shaft position is required to be compared with an input shaft position. For this purpose, two or

more rotary transformers are used in a subsystem or in conjunction with the servo system. In such cases, one or more of the devices may be used to receive, modulate, or transmit data while another may be used to produce a torque that is a function of the input received from the data-handling units. For the present discussion, the torques produced by rotary transformers will be considered to be in the order of a few inch-ounces because, in a strict sense, rotary transformers are small-torque devices.

11-3.4 Miscellaneous System Uses

Rotary transformers are also used in systems that require modulation of electrical waveforms, resolution of vectors into components, and vector combining processes. In general, such uses are concerned only with data-handling operations.

11-3.5 General Functional Classification

Considering the above, rotary transformers may be classified according to their use in systems, such as: data-handling units; and torque-producing units. The category into which a unit falls depends upon the load requirements. If the required torque output is furnished by the device handling the data, the device is classified as a data-handling unit. On the other hand, if the required torque output exceeds the capacity of the data-handling device, a torque-producing unit must be used.

*By A. Kusko

11-3.6 General Unit Classification

Other general classifications of rotary transformers involve their construction, basic principles of operation, or a combination of these categories. This method is used to designate the general type of rotary transformers discussed in the following sections; i.e., synchros, induction resolvers, induction potentiometers, toroid-wound rotary transformers, and microsyns. Further classifications of these general types are pointed out as they occur.

11-3.7 SYNCHROS

Synchros are rotary transformers that, with minor changes in basic design, may be used for either data-handling or torque-producing functions. Most synchros have a configuration similar to that of a conventional 3-phase alternator of fractional horsepower size. The form of the rotor and the arrangement of the rotor winding identify the type of synchro and its function.

11-3.8 Stator Construction

In general, the synchro stator is a cylindrical slotted and laminated structure with three windings arranged in the slots at 120° spatial displacement from each other. In most units, the slots are skewed one slot pitch to avoid "slot lock" and resulting angular errors. Units that do not have skewed stator laminations are constructed with skewed rotor laminations.

The stator windings are not 3-phase in the usual sense because all induced voltages are in time phase. They can be either Y-connected or delta-connected (Fig. 11-17) and serve as the secondary winding of the synchro. Stator connections are usually brought out as three leads and labelled S_1 , S_2 , and S_3 .

11-3.9 Rotor Construction

Rotors of standard synchros are of two-pole construction, the most common type being the salient-pole rotor used in transmitter or repeater units. The rotor is a slotted and laminated structure of the "dumbbell" or "H" type. The structure is mounted on a shaft that turns on ball bearings. It carries a

machine-wound single-phase spool winding that serves as the primary winding of the synchro. Connections are made to the primary winding through two slip rings and brushes that have full excitation voltage impressed upon them at all times. Rotor lead connections are usually labelled R_1 and R_2 .

11-3.10 Synchro Supply

All standard synchros are designed to operate from one of the following supplies: (a) 26 volts at 400 cps, (b) 115 volts at 400 cps, or (c) 115 volts at 60 cps.

11-3.11 Nomenclature

Military Specification MIL-S-20708 describes the method of classification to be used for standard synchros of new design. The type designation of a standard synchro identifies its voltage, size, function, supply frequency, and modification as follows:

(a) Both 26-volt and 115-volt synchros are classified in the same manner, except that the type designation of 26-volt synchros is prefixed by "26V"

(b) The first two digits indicate the *maximum diameter* in tenths of an inch. If the diameter is not a whole number of tenths, the next higher tenth is used.

(c) The succeeding group of letters indicates the *function* in accordance with the following:

| <i>First Letter</i> | <i>Function</i> |
|---------------------|-----------------|
| C | Control |
| T | Torque |

| <i>Succeeding Letters</i> | <i>Function</i> |
|---------------------------|--------------------------|
| D | Differential |
| R | Receiver |
| T | Transformer |
| X | Transmitter |
| B | Rotatable Stator Winding |

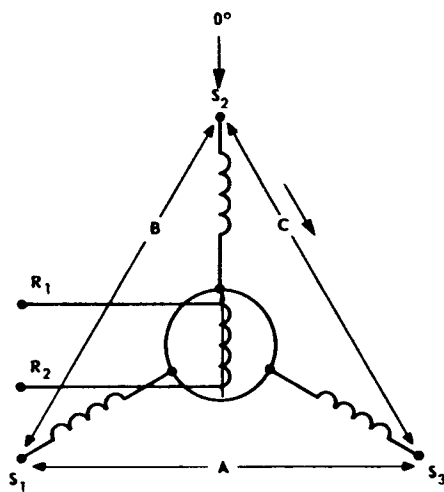
When two synchros are enclosed within the same housing, the type designation indicates both units, e.g., 37 TR-TR6a.

SENSING ELEMENTS

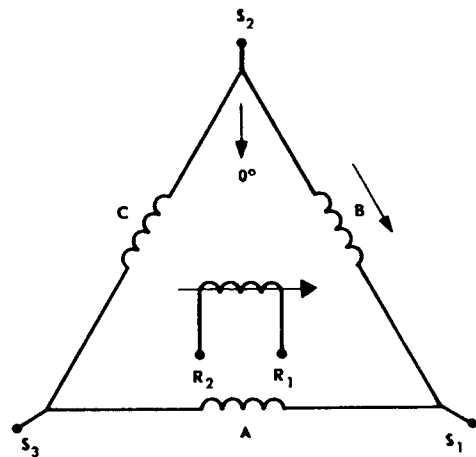
(d) The succeeding digit indicates the *frequency* of the power supply in accordance with the following:

| Number | Supply Frequency (cps) |
|--------|------------------------|
| 6 | 60 |
| 4 | 400 |

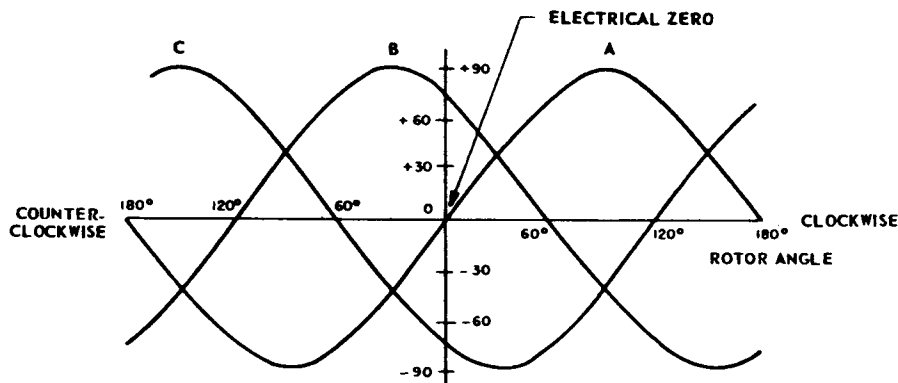
(e) The lower-case letter "a" following the frequency digit indicates the original issue of a standard synchro type. The first *modification* that affects the external mechanical dimensions or the electrical characteristics of the basic type is indicated by the letter "b", succeeding modifications being indicated by "c", "d", etc.



A. Y-CONNECTED STATOR



B. DELTA-CONNECTED STATOR



C. RMS STATOR VOLTAGES VS ROTOR ANGLE

Fig. 11-17 Electrical characteristics of a typical synchro.

MEASUREMENT AND SIGNAL CONVERTERS

As an illustration of the interpretation of the type designation, an 18CDX4b synchro is the first modification of a 115-volt, 400-cycle, control differential transmitter whose largest diameter is between 1.71 and 1.80 inches. If it were a 26-volt synchro, its designation would be 26V-18CDX4b.

11-3.12 Other Methods of Nomenclature

The various services formerly used individual methods of nomenclature which will still be encountered in old designs. The Army method of nomenclature is much like the system described in the previous paragraph, while the Navy method for 115-volt 60-cps synchros differs greatly. The latter synchros have a code designation such as 5 HCT Mark 2 Mod 3B, which identifies the approximate size, special design features, function, type designation, and manufacturer as follows:

(a) The first digit indicates a size group in the following table:

| Size | Approx Wt (lb) | Approx Length (in.) | Approx Dia (in.) |
|------|----------------|---------------------|------------------|
| 1 | 2 | 3.9 - 4.2 | 2.25 |
| 3 | 3 | 5.2 - 5.51 | 3.1 |
| 5 | 5 | 6.0 - 6.8 | 3.39 - 3.625 |
| 6 | 8 | 6.4 - 7.5 | 4.5 |
| 7 | 18 | 8.9 - 9.2 | 5.75 |

Other synchros such as Army units or types with trade names are often referred to as "Size 1" etc., if they approximate the corresponding Navy type.

(b) One or more modifying letters following the first digit indicate that the unit includes *special fittings* as follows:

B — bearing-mounted

F — flange-mounted; this letter is omitted if letters other than "H" and "S" occur in a receiver type designation

H — special bearings and brushes for high-speed operation (1200 rpm for 1500 hours)

N — nozzle-mounted

S — special unit; i.e., synchro does not conform to standard specifications and therefore is not interchangeable with other units having the same designation but different "Mod" number

(c) One or more letters following the special-fitting letters indicate the function of the unit from Table 11-4.

(d) The *Mark number* signifies the design of the particular unit being specified.

(e) The *Mod number* designates the manufacturer as assigned by the Bureau of Ordnance.

11-3.13 Transmitter Characteristics

The transmitter generates and transmits signal voltages with a magnitude corresponding to the angular position of the rotor. Consider the rotor (or primary) winding to be supplied with a single-phase a-c excitation current. This excitation sets up a flux that varies sinusoidally with the excitation frequency and links each of the three stator (or secondary) windings. The extent of the linkage depends upon the angular position of the rotor, the distribution (120° spatial displacement) of the stator windings, and the configuration of the rotor and stator pole faces. For a given unit, therefore, the magnitude of the single-phase voltage induced in each of the stator windings is a function of the angular position of the rotor.

A plot of the induced stator voltages for different rotor positions is shown in Fig. 11-17. The curves in this plot apply to either Y-connected or delta-connected transmitters. This result makes it difficult to tell whether the stator of a given machine is Y-connected or delta-connected because either type of connection yields exactly the same stator voltages for each position of the rotor.

It follows from the above that, for a given distribution and polarity of the stator voltages, there is but one corresponding rotor position. Conversely, for a given position of the

SENSING ELEMENTS

TABLE 11-4 FUNCTIONAL CLASSIFICATION OF SYNCHROS

| Functional Designation | Symbol | Function |
|----------------------------------|--------|--|
| Synchro transmitter | X | Generates and transmits electrical signal corresponding to the angular position of the mechanically driven rotor of the unit. |
| Synchro receiver | R | Produces a torque through the turning of the rotor shaft as a function of an electrical input signal received from the transmitter. |
| Synchro control transformer | CT | Produces an electrical output signal having a magnitude proportional to the angle of rotation of the unit rotor with respect to the magnetic field set up by the unit stator. Indirectly, the output signal is a sine function of the angle between the rotor shaft position and the shaft position of the transmitter that generates the transformer input. |
| Synchro differential transmitter | DT | Modifies a received signal and transmits an electrical signal corresponding to the sum or difference of the impressed and modifying signals. The modifying signal is proportional to the angle of rotation of the mechanically driven rotor shaft. |
| Synchro differential receiver | DR | Produces a mechanical output signal when the rotor shaft of the unit turns in accordance with the sum or difference of electrical signals received from two sources. |

rotor, there is but one stator voltage condition. The electrical zero position for Y-connected and delta-connected stators is shown in Fig. 11-17.

11-3.14 Receiver Characteristics

A receiver has the same electrical configuration as a transmitter but the unit output is an angular position of the rotor corresponding to an electrical signal input. The rotor winding of the receiver is excited from the same single-phase a-c source as the transmitter rotor, and the receiver stator terminals are connected to corresponding terminals of the transmitter (Fig. 11-18). Assume for the moment that the rotor winding of the receiver is open and that voltages are induced

in the stator windings of the transmitter. These voltages are applied to the stator windings of the receiver and the resulting current flow sets up a magnetic flux in the receiver air gap. Moreover, because the transmitter and the receiver windings have the same configuration and voltages, the orientation of the magnetic field in the receiver is identical with that in the transmitter. Now assume that the receiver rotor winding is switched to the same a-c excitation source that supplies the transmitter rotor. This operation sets up a second polarized field in the receiver air gap. As the second field tends to align itself with the magnetic field induced by currents in the stator windings, the rotor turns until the two

fields are aligned. The receiver rotor, therefore, is forced to take up an angular position corresponding to the position of the transmitter rotor.

A distinguishing feature of a receiver is an oscillation damper in the form of a flywheel that is free to rotate relative to the rotor shaft within limit stops set about 45° apart. The flywheel has approximately the same moment of inertia as the rotor. A friction coupling between the rotor shaft and the flywheel serves to dissipate energy when the rotor oscillates as it does when coming into a position of alignment. The added inertia furnished by the oscillation damper also prevents the rotor from "running away" when it is forced to follow the transmitter rotor in a continual process. Without the oscillation damper, "running away" would occur at high angular velocities when the torque becomes large enough to overcome the "aligning" torque and thereby prevents the transmitter from maintaining control.

Receivers excited from a 60-cps source can be used as transmitters but the converse is not true. In general, receivers designed for 400-cps application have little tendency to run away and therefore are not provided with dampers.

11-3.15 Transformer Characteristics

A transformer is used with a transmitter to indicate the difference between the angular positions of the shafts of each unit. The electrical connections for this purpose are shown in Fig. 11-19. The synchro capacitors included in this figure are discussed in Par. 11-3.18.

The stator windings of a transformer are similar to those of a transmitter except that each winding of the transformer has more turns of wire and therefore a higher impedance. This characteristic reduces the excitation current drawn from the system by the addition of the transformer. A transformer rotor differs from a transmitter rotor in that

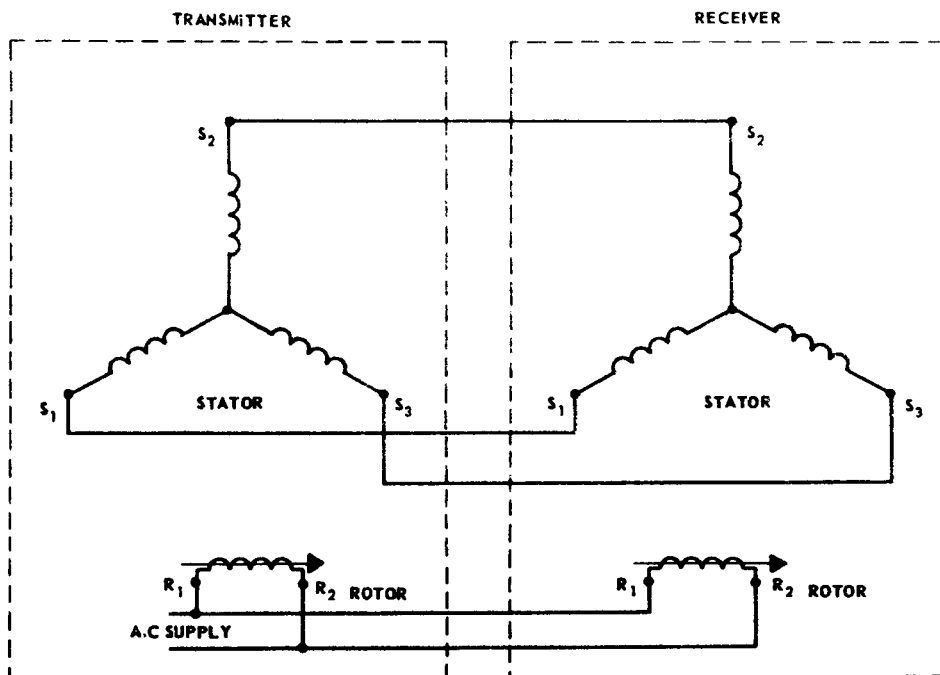


Fig. 11-18 Electrical connections for transmitter-receiver system.

SENSING ELEMENTS

it has a cylindrical instead of a salient-pole shape and a single high-impedance winding with a large number of turns of fine wire. The cylindrical rotor provides for a constant input impedance regardless of the rotor angular position. The high impedance of the rotor winding provides for a higher and more useful output-voltage gradient

In a transmitter-transformer system, the transmitter stator voltages cause excitation currents to flow in the transformer stator windings. These currents produce a flux corresponding to the angular position of the transmitter rotor. The voltage induced by the flux in the transformer rotor has a magnitude that depends upon the position of the transformer rotor with respect to the flux and therefore with respect to the position of the transmitter rotor. It follows that the induced voltage is at a maximum when the transformer rotor is in a position to link the maximum flux. Hence, a 90° displacement of the rotor

from its maximum position places the rotor coil across the flux and the induced voltage is zero. The latter position (often called the "null" position) is the normal condition of transformer operation.

From the above, the transformer output voltage is a function of the error angle between the relative rotor positions of the transmitter and the transformer. Moreover, by transformer design, the magnitude of the output voltage is proportional to the sine of this error angle.

The impedances of the stator and rotor windings of a transformer are considerably higher than those of an equivalent-sized transmitter or receiver. Hence, a transformer should never be used to feed a low-impedance load. Because the output of a transformer is normally connected to an amplifier with high input impedance, the rotor current and therefore the developed torque are negligible.

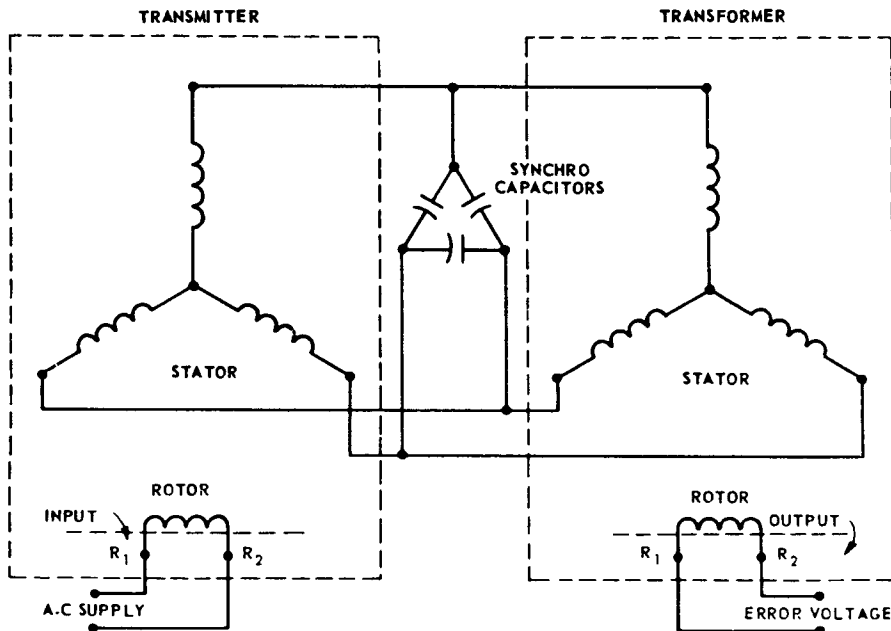


Fig. 11-19 Electrical connections for transmitter-transformer system.

11-3.16 Differential Transmitter Characteristics

A differential transmitter is commonly used as a component of a differential synchro system. Such a system is used when it is desired to have the angular position of an output shaft equal the sum or difference of the angular positions of two input shafts.

A differential transmitter has a stator identical with that of a transmitter or receiver but the rotor is cylindrical instead of salient-pole and has three windings spaced 120° apart like the stator. The turns ratio between the rotor and stator windings is approximately 1:1. Three slip rings and brushes are used for connections to the rotor windings (Fig. 11-20). Both Y-connections and delta-connections may be used. Normally, the differential transmitter is an intermediate unit connected between the output of a transmitter and the input of a receiver or a transformer. By this arrangement, the differential transmitter receives an electrical signal corresponding to the angular position of the transmitter rotor, modifies this electrical signal by an amount corresponding to its own rotor position, and transmits the modified electrical signal to the stator of the receiver or the transformer. This electrical input to the receiver creates a magnetic field having an orientation that is either the sum or difference of the rotor angles (or inputs) of the transmitter and the differential transmitter.

The output signal of the differential transmitter system equals the sum or difference of the two input signals depending upon the following:

(a) The relative directions of rotation of the transmitter rotor and the differential transmitter rotor.

(b) The electrical connections between the transmitter and the differential transmitter.

(c) The electrical connections between the differential transmitter and the receiver or transformer.

Different relative directions of rotation can be obtained by interchanging pairs of the input and the output leads of the differential transmitter. An arrangement of leads as

shown in Fig. 11-20 is called a *symmetrical* connection. The receiver angle for the symmetrical connection, and changes thereto, equals the sum or difference of the two input angles as follows:

(a) Symmetrical connection (Fig. 11-20):
 receiver angle = transmitter angle
 — differential transmitter angle

(b) Two input and output leads interchanged: (e.g., transmitter S_1 and S_3 to differential transmitter S_3 and S_1 , respectively, and differential transmitter R_1 and R_3 to receiver S_3 and S_1 , respectively)

receiver angle = transmitter angle
 + differential transmitter angle

(c) Two input leads interchanged: (e.g., transmitter S_1 and S_3 to differential transmitter S_3 and S_1 , respectively)

receiver angle = — (transmitter angle
 + differential transmitter angle)

(d) Two output leads interchanged: (e.g., differential transmitter R_1 and R_3 to receiver S_3 and S_1 , respectively)

receiver angle = differential transmitter
 angle — transmitter angle

When the rotor windings of the differential transmitter in Fig. 11-20 are connected to a transformer instead of a receiver, the transformer output voltage is a measure of the error angle between the transformer rotor angle and the combined rotor angle of the transmitter and differential transmitter that positions the transformer flux.

11-3.17 Differential Receiver Characteristics

Differential synchro systems also employ a differential receiver as an intermediate unit between two transmitters (see Fig. 11-21). The construction of a differential receiver is similar to that of a differential transmitter except that the differential receiver, as an ordinary receiver, has an oscillation damper. In a differential receiver system, the stator of the differential receiver is connected to the stator of one of the associated transmitters and its rotor is connected to the stator of the

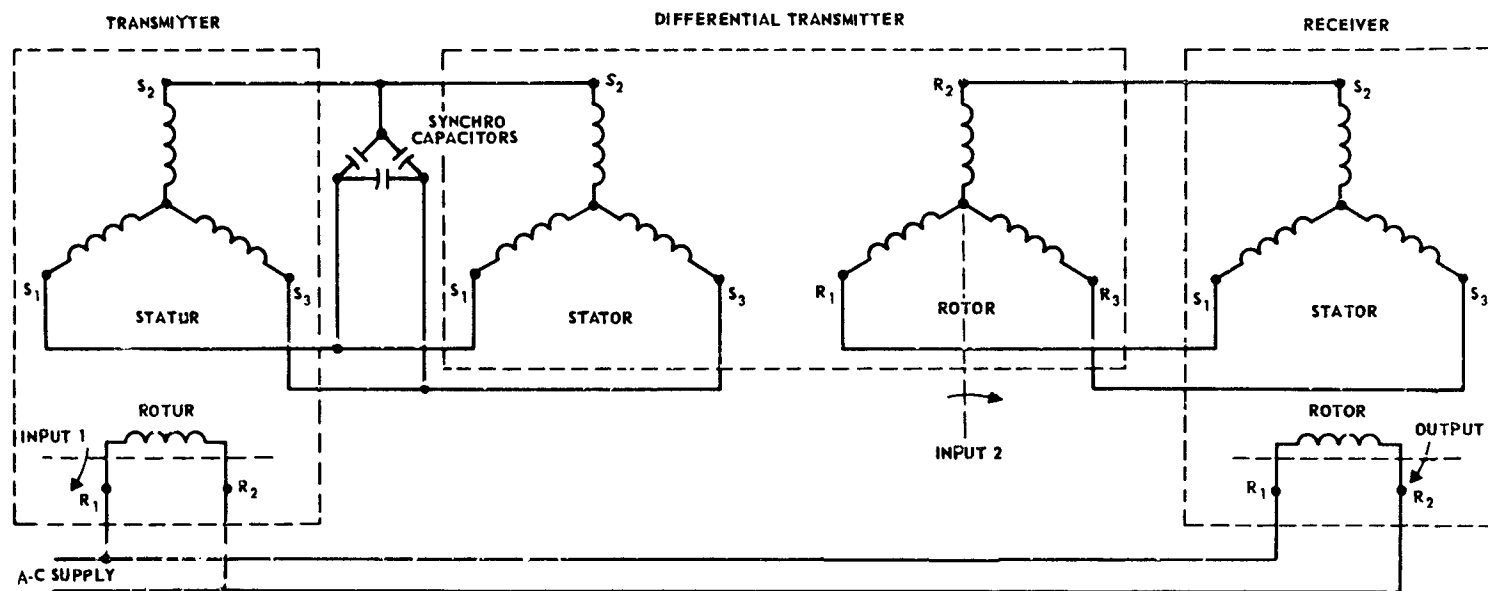


Fig. 11-20 Electrical connections for a differential transmitter system.

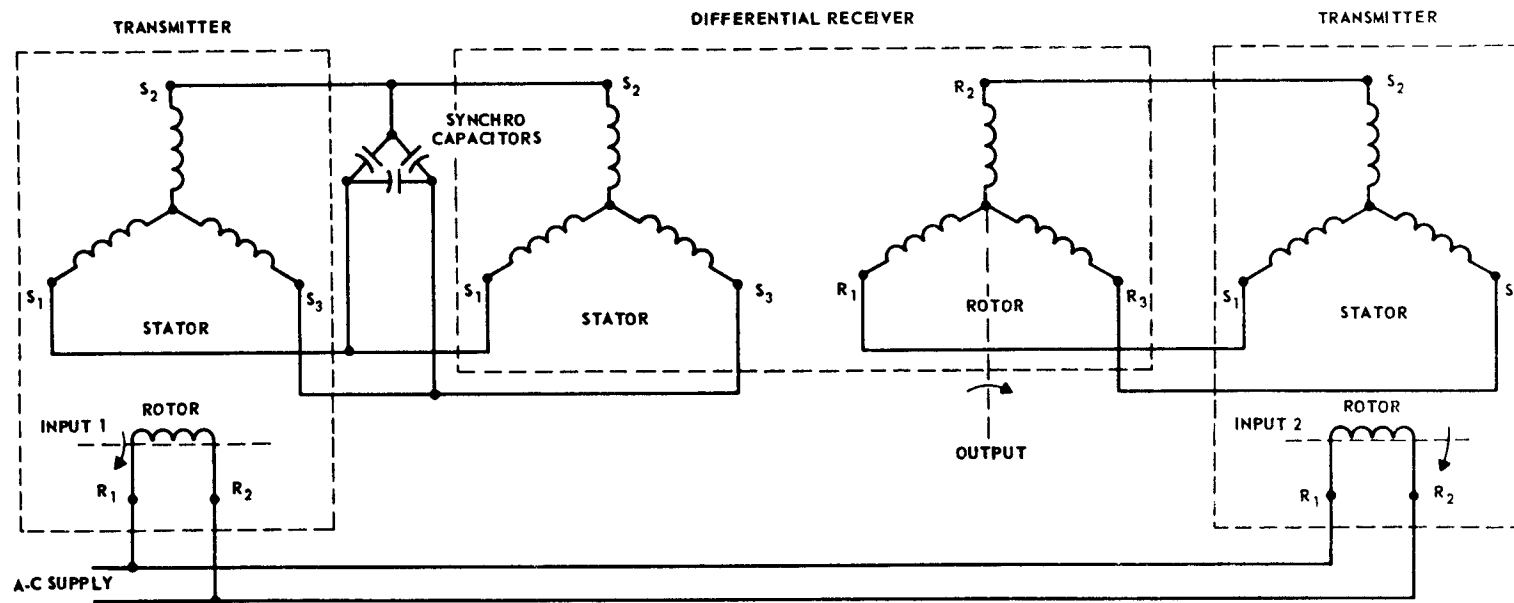


Fig. 11-21 Electrical connections for a differential receiver system.

other transmitter. The differential receiver therefore has two electrical inputs and its shaft output is the mechanical output of the system.

The voltages applied to the stator and to the rotor of the differential receiver depend upon the rotor positions of the supplying transmitters. It follows that two magnetic fields are created in the gap of the differential receiver, one depending upon the rotor position of one transmitter and the other depending upon the rotor position of the other transmitter. Hence, if the rotor of the differential receiver is free to turn, it assumes a position in which the flux vectors of both magnetic fields are aligned. In this position, the rotor angle of a differential receiver connected as in Fig. 11-21 equals the rotor angle of the left transmitter minus the rotor angle of the right transmitter. Reversing pairs of connections at the differential receiver changes the relative directions of motion on the two transmitter shafts as previously described for the differential transmitter.

11-3.18 Synchro Capacitors

Synchro capacitors are used with transformers, differential transmitters, and differential receivers to neutralize the lagging component of the magnetizing current drawn from the transmitters supplying the stators of the units concerned. For this purpose, three matched capacitors are delta-connected as shown in Figs. 11-19 to 11-21. The leading current drawn by the capacitors at almost zero power factor neutralizes the lagging component of the magnetizing current and thus reduces the total current drawn from the transmitter. To ensure proper operation, the capacitance of the three legs of the capacitor must be balanced to within 1 percent but the absolute value of the legs may vary over ± 10 percent. For maximum effectiveness, the leads connecting the capacitor bank to its associated transformer or differential unit should be as short as possible.

11-3.19 Dual-Speed Synchro Systems

To provide synchro systems with a greater over-all accuracy, it is possible to gear up the

transmitter input shafts from their driving devices and to gear down the shaft output of the associated receiver or transformer by the same ratio. Such gearing reduces the effect of the synchro no-load errors by a factor equal to the gear ratio. Load errors are reduced even more because the torque at the load shaft causes less torque on the receiver and a smaller error at the receiver shaft. However, the friction and backlash of the gears introduce errors; and a geared system is not completely self-synchronous because it is possible for the receiver to "slip a pole" and find a synchronous position other than the correct one.

Consider a 36-speed system where one revolution of the driving shaft turns the transmitter rotor through 36 revolutions, and 36 revolutions of the transformer or receiver shaft correspond to 1 revolution of the system output shaft. In such a system, there are 35 incorrect positions of the input-output relationship for one correct position. To obtain accuracy without such false indications, it is common practice to add a duplicate system of synchros operating at a low ratio (usually 1:1). The combination of the 36-speed and the 1-speed systems is referred to as a *dual-speed* system. Incorrect alignment of the 36-speed system is prevented because, when the error is large, the 1-speed system has control and reduces the error to a small value. The 36-speed system then takes control and maintains its inherent high accuracy.

The use of speed in connection with the system as a whole should not be confused with the gear ratio; e.g., an n -speed synchro is one that turns n times for one turn of the actuating device. Common speeds are 1, 2, 10, 18, and 36. Note also that, in a dual-speed system, all elements in the system must be duplicated. For example, in Fig. 11-21, two pairs of transmitters and also two differential receivers would be required.

11-3.20 Zeroing

For correspondence between the angular indications at the transmitting and receiving ends of a synchro system, it is necessary to orient the units at each end with respect to

each other. For this purpose, a definite relative position of the rotor with respect to the stator has been designated as the "electrical zero" position for all standard synchro units. This position is the relative angular position of the rotor and the stator that results in a minimum, or null, output voltage. The process of rotating the stator or rotor with respect to one another to obtain the "electrical zero" point is called *zeroing*.

To facilitate the zeroing of synchros, the units are constructed with standardized flanges concentric with the shaft. In some synchros, the flanges are located on the head end of the frame; in others, near the center. In either case, the synchros are mounted and held in place by a ring retainer or by dogs acting on the flange. Thus, by loosening the ring or the dogs, the stator may be rotated until the synchro output corresponds to the zero shaft position. When this procedure is followed for all synchros in a system, the system will operate properly.

In Fig. 11-17, it is shown that the voltage between stator terminals S_1 and S_3 is zero at zero rotor angle. However, this one condition is not sufficient to specify "electrical zero" because the $S_1 - S_3$ voltage is also zero at a 180° rotor angle. For zeroing purposes, therefore, use is made of the fact that the effective voltages of the stator and rotor windings are in the same direction at "electrical zero."

11-3.21 Zeroing a Transmitter (or Receiver)

The zeroing of a transmitter or receiver may be accomplished with the connections shown in Fig. 11-22. A supply of rated voltage and frequency is used. In determining the approximate zero, the shaft must be free to rotate when power is applied to prevent damage to the synchro.

Using the connections shown in Fig. 11-22A, the rotor will turn by itself to an approximate zero position. An accurate zero is then obtained with the stator terminals open and the voltmeter connected across S_1 and S_3 (Fig. 11-22B). With this arrangement, the rotor or stator is turned slowly until the voltage $S_1 - S_3$ is a minimum.

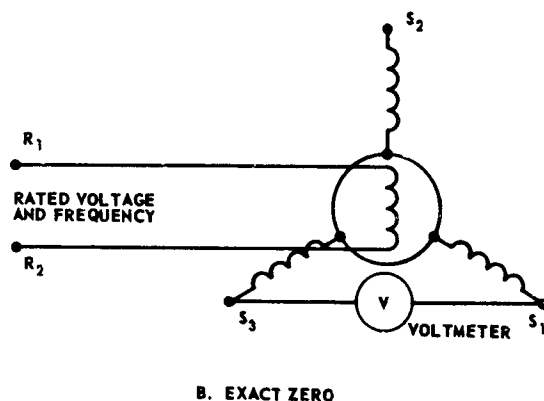
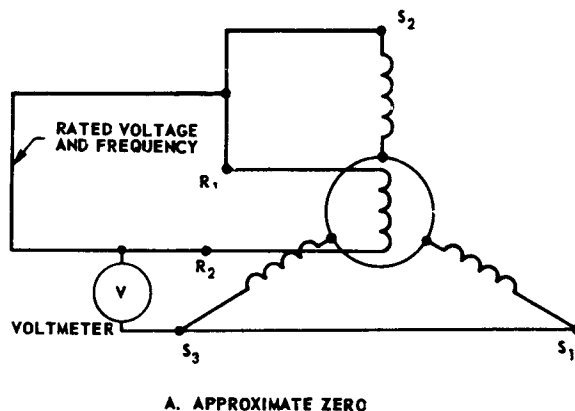


Fig. 11-22 Connections for zeroing a transmitter or receiver.

11-3.22 Zeroing a Control Transformer

An approximate zero position of a control transformer is obtained with the connections shown in Fig. 11-23A. Approximate zero is the position obtained by rotating the rotor or stator until the voltage reading is a minimum. Then, using the connections of Fig. 11-23B, a fine adjustment to a zero voltage reading will provide an accurate zero setting.

11-3.23 Zeroing a Differential Synchro Unit

Using the connections shown in Fig. 11-24A, the rotor of a differential synchro unit will turn by itself to an approximate zero position. For an accurate zero determination, connect the unit as shown in Fig. 11-24B and adjust until the voltage across R_1 and R_3 is zero.

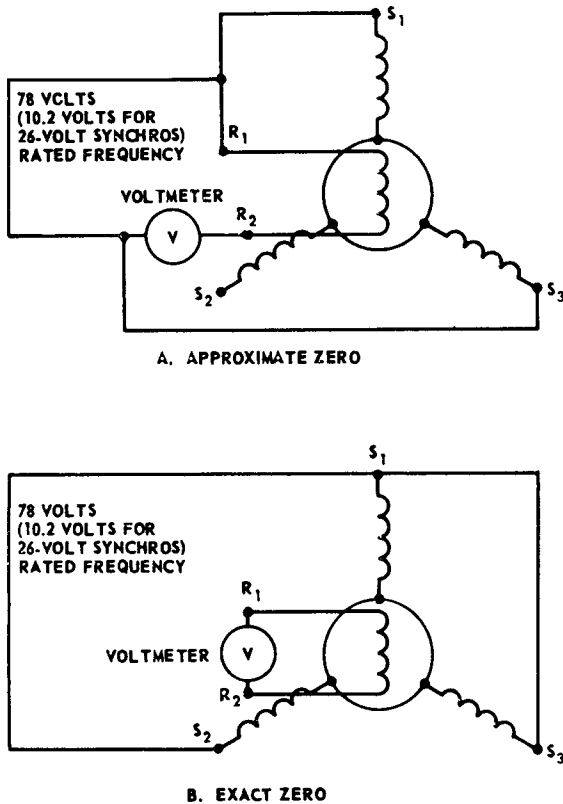


Fig. 11-23 Connections for zeroing a control transformer.

11-3.24 Torque Relationships

Many synchro applications are based on a continuous and accurate alignment between the transmitter and receiver shafts. For this purpose, the receiver must generate a torque to provide shaft motion and acceleration because the receiver plus the load connected to it always have some friction and inertia. This fact makes perfect alignment impossible because no torque is developed when the two rotor positions coincide.

The torque or rotor current as a function of the displacement angle is shown in Fig. 11-25. Note that the rotor current varies with angular displacement and has its minimum or null point at zero displacement. Note also that the torque is approximately a sinusoidal

function of the angle between the transmitter and receiver shafts. The maximum torque occurs at 90° and this torque determines the maximum load that can be handled by the receiver. For accuracy, the receiver should be operated at light loads and over small continuous angular displacements (e.g., not exceeding 20°) so that momentary overloads will not pull the motor out of step. Operation at light loads also prevents overheating of the units.

11-3.25 Torque gradient. The torque gradient of a synchro is defined as the slope of the torque curve at the point of zero displacement (Fig. 11-25). This gradient is expressed in inch-ounces per degree of rotor displacement. Because an increase in the torque gradient gives a proportional decrease in the static error, it is desirable that the receiver have a large torque gradient.

11-3.26 Performance prediction. At times, it is helpful to predict the performance of a synchro system in which the receiver is connected to a transmitter of different size and construction. The method used for such a prediction is based on the following known operating characteristics:

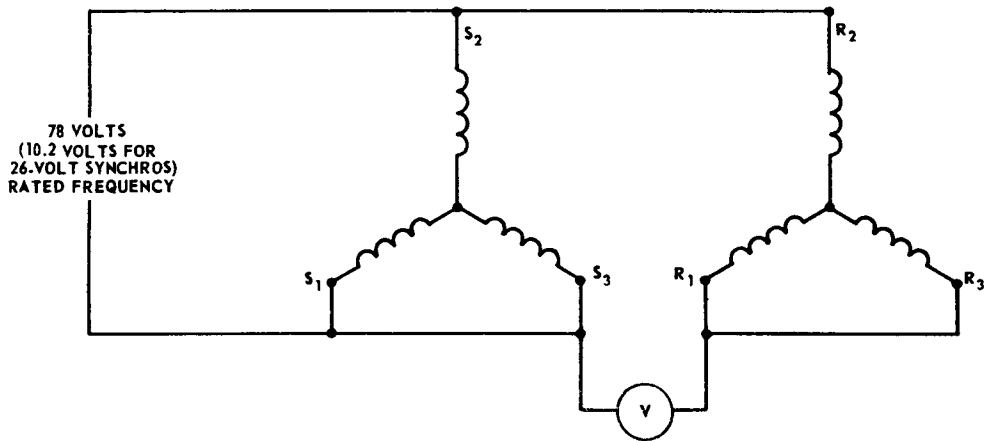
(a) The unbalanced stator voltages of a standard transmitter or receiver are the same regardless of the size of the units.

(b) The only electrical characteristic that varies with the size of a synchro is the internal impedance of the stator windings.

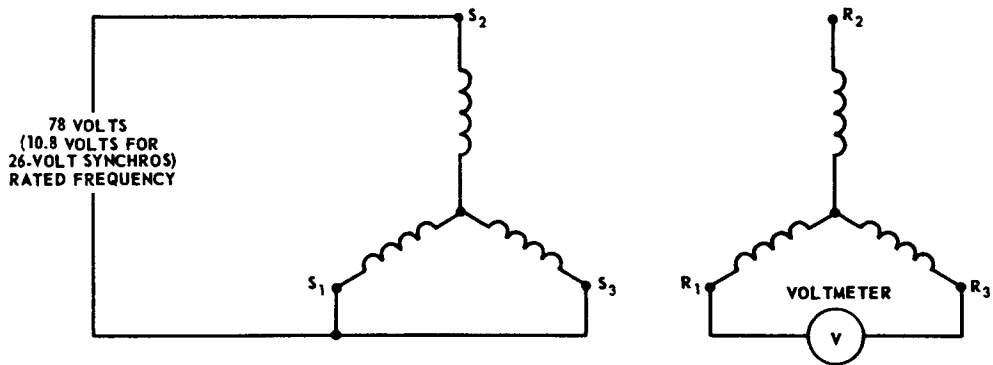
(c) When two standard synchros of the same size are connected together, the internal impedance determines the amount of current flowing and thereby the torque produced.

(d) The torque produced by the synchros in (c) can also be determined by measuring the torque gradient of the receiver. A torque gradient obtained in this manner is called the *unit torque gradient* and is inversely proportional to the internal impedance of the stator windings.

In view of the above, consider a receiver driven by a transmitter having a unit torque gradient equal to R times that of the receiver.



A. APPROXIMATE ZERO



B. EXACT ZERO

Fig. 11-24 Connections for zeroing a differential synchro unit.

Then the actual receiver torque gradient will be $2R/(1 + R)$ times the receiver unit torque gradient. Consider now that a large transmitter is used to drive a number (N) of small receivers connected in parallel, each having an equal load. Under these conditions, each receiver will develop an actual torque gradient of $2R/(N + R)$ times its unit torque gradient. This relationship does not apply when the receivers are loaded unequally or in systems using differential units or control transformers.

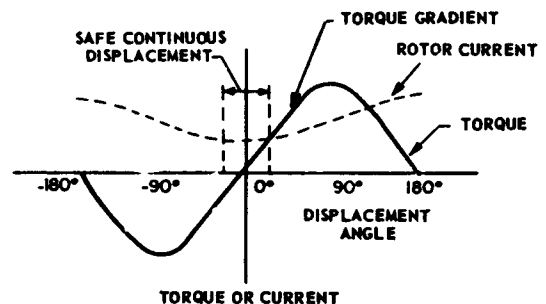


Fig. 11-25 Torque and rotor current as a function of displacement angle.

11-3.27 Synchro Accuracy

Under static conditions, all synchros and synchro systems are subject to errors created by unavoidable imperfections in design, manufacture, or operation. Torque-producing synchros are also subject to dynamic errors introduced while overcoming friction and inertia. For specified operating conditions, the maximum total error of typical data-handling synchros (transmitters, differential transmitter, and control transformers) corresponds to an angle in the order of 10 minutes of arc. For torque-producing synchros (receivers and differential receivers) the maximum total error is in the order of 1.5° . These maximum error values may be increased or decreased slightly depending upon the design of the unit, its manufacture, and its arrangement in a system.

As mentioned previously, perfect alignment between the transmitter and receiver shafts at the null point is an impossibility. This case is an example of an unavoidable design imperfection. An example showing how synchro accuracy can be improved by proper design is explained in Par. 11-3.18 on synchro capacitors. Manufacturing imperfections that produce static errors are as follows:

- (a) Irregularities of the stator bore
- (b) Rotor eccentricity

(c) A nonuniform air gap resulting from slot openings around the stator periphery.

The above imperfections introduce electrical errors as shown in Fig. 11-26. Another source of static error is electrical noise.

11-3.28 Static errors. Static errors tend to increase the null voltage. The amount of null voltage that can be tolerated depends upon the amplifier gain and the ability of the amplifier to distinguish between the quadrature and the in-phase signal. The maximum fundamental component of null voltage in typical synchros is 30 millivolts. Depending upon the type of synchro, service specifications usually restrict the value of null voltage to a range of 40 to 60 rms millivolts.

Because an increase in torque gradient gives an approximately proportional decrease

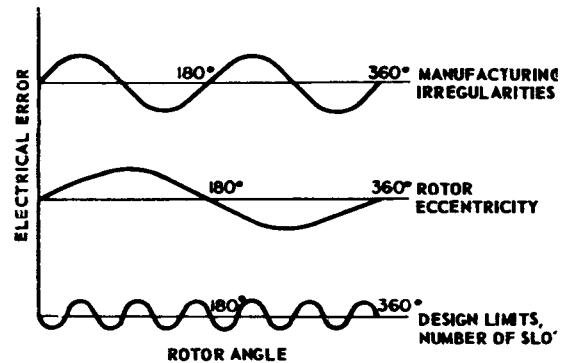


Fig. 11-26 Sources of error in a synchro unit.

in static error, the error of a transmitter-receiver system can be predicted by the method used previously for determining the actual torque gradient of a receiver. The static error of a transmitter-control transformer system may be predicted approximately from the no-load electrical errors of the individual units. If the units are connected so that their shafts turn in the same direction, the no-load electrical error of the system equals the electrical error of the control transformer minus the electrical error of the transmitter. Hence, system accuracy improves if the no-load electrical error curves of both units follow approximately similar variations of the shaft position.

The above methods cannot be used to determine the over-all error of a system employing a differential synchro because the electrical error of such a unit is a function of the rotor positions of other units connected to it. In general, the over-all error of a differential synchro system is the sum of the errors of the individual units.⁽²²⁾

11-3.29 Dynamic errors. The dynamic error of a receiver and differential receiver is the angle by which the receiver shaft lags behind the transmitter shaft during a slow rotation of the transmitter shaft. The chief sources of this type of error are the friction of the receiver bearings and brushes and the inertia of

MEASUREMENT AND SIGNAL CONVERTERS

the receiver rotor and load. Another type of dynamic is called a *speed* error because it is a function of the synchro operating speed. The speed error is explained by examining the voltage induced in the stator windings of a transmitter. This voltage is the result of a time rate of change of the flux set up by the currents in the rotor winding. The stator flux ϕ_s linked by a stator coil, the axis of which is inclined to the rotor axis at an angle θ , may be expressed as

$$\phi_s = KE \cos \omega t \sin \theta \quad (11-24)$$

where

E = applied rotor voltage

K = winding and frequency constant for the synchro

$\omega = 2\pi f$ = radian frequency of the voltage applied to the rotor

Thus, the voltage induced in a stator coil of N turns is

$$e = N \frac{d\phi_s}{dt} = NKE \cos \omega t \cos \theta \frac{d\theta}{dt} - NKE\omega \sin \omega t \sin \theta \quad (11-25)$$

The desired induced voltage is in the form

$$e = K' \sin \theta \sin \omega t \quad (11-26)$$

Identification of the constant K' in Eq. (11-26) with the factor $-NKE\omega$ in Eq. (11-25) shows

that the term $NKE \cos \omega t \cos \theta \frac{d\theta}{dt}$ in Eq.

(11-25) is an additional or error voltage. This error voltage is present only while the rotor of the transmitter is rotating, since it is

proportional to $\frac{d\theta}{dt}$. If the transmitter is

driven at constant speed, $\frac{d\theta}{dt}$ is a constant,

**TABLE 11-5 LIMITS OF TOLERANCES FOR MILITARY SYNCHROS
(115 VOLTS, 60 CPS)**

| Type of Synchro | Transmitter | | | | | Receiver | | |
|---|-------------------|-------------------|-------------------|-------------------|--------------------|----------------------|------------------------|------------------------|
| Size | 1 G or 1 HG | 3 G or 3 HG | 5 G or 5 HG | 6 G or 6 HG | 7 HG or 7 HG | 1 HF 1 F | 3 F 3 HF 3 B | 5 F 5 HF 5 B |
| Static accuracy max error (minutes) | 18 | 18 | 18 | 18 | 18 | 90(1 F) 150(1 HF) | 36(3F, 3B) 60(3 HF) | 36(5F, 5B) 45(5 HF) |
| Torque gradient min (oz-in./deg) | 0.06 | 0.25 | 0.40 | 1.2 | 3.4 | 0.06 | 0.25 | 0.40 |
| Secondary peak- voltage limits | | | | | | | | |
| min (volts) | 88.2 | 88.2 | 88.2 | 88.2 | 88.2 | 88.2 | 88.2 | 88.2 |
| max (volts) | 91.8 | 91.8 | 91.8 | 91.8 | 91.8 | 91.8 | 91.8 | 91.8 |
| Primary current max (amp) | 0.30 | 0.40 | 0.60 | 1.3 | 3.0 | 0.30 | 0.40 | 0.60 |
| Primary power max (watts) | 4.8 | 5.5 | 7.0 | 15.0 | 22.0 | 4.8 | 5.5 | 7.0 |
| Secondary load current | | | | | | | | |
| max (amp) | 0.04 | 0.20 | 0.35 | 0.70 | 1.50 | 0.04 | 0.20 | 0.35 |
| Temperature rise max (°C) | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |

CHAPTER 12

SIGNAL CONVERTERS*

12-1 INTRODUCTION

12-1.1 TYPES

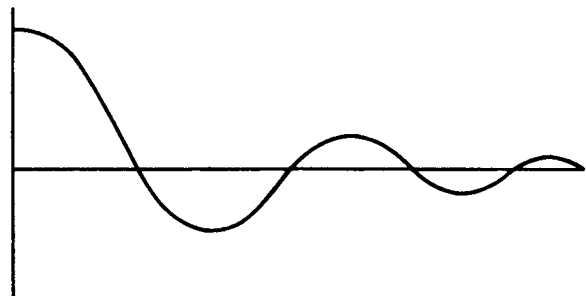
Three types of signal converters are discussed in this chapter: modulators, which superimpose a signal on a carrier; demodulators, which recover the signal that has been superimposed on a carrier; and digital-to-analog converters, which express a number in electrical form as a proportionate voltage or as a shaft rotation.

12-1.2 Modulators

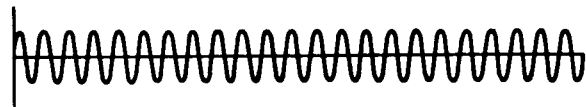
Modulators are used in servomechanisms in which the error signal is a direct voltage, but the output member, such as a 2-phase motor, will respond only to a-c signals. The modulator converts the d-c signal to a-c. Another use for modulators in servomechanisms is the case where it is undesirable to employ d-c amplification of the d-c error signal because of drift problems in such an amplifier. Here, to avoid the drift problems, the signal is first modulated, then amplified by an a-c amplifier, which is drift free; the amplifier output is then demodulated to recover the d-c signal for application to the output member.

12-1.3 Demodulators

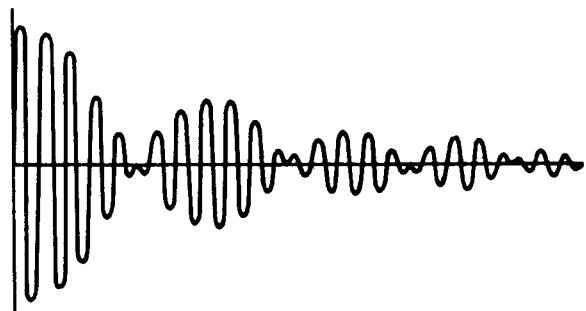
Demodulators are used in servomechanisms in which the error signal is an a-c voltage (such as the output of a synchro control transformer) and the output member responds only to d-c signals (e.g., an amplidyne).



A. ERROR SIGNAL



B. CARRIER SIGNAL



C. MODULATOR OUTPUT

Fig. 12-1 Modulator waveforms.

*By A. K. Susskind

12-1.4 Form of Modulation

The form of modulation used in servomechanisms is called *suppressed-carrier* modulation. This means that the output of the modulator contains frequency components which are the sum and difference of the carrier and the signal frequencies, but contains no frequency component equal to the carrier frequency. Figure 12-1A shows an error signal and Fig. 12-1C shows the corresponding output of a suppressed-carrier modulator. The modulated signal is zero when the error signal is zero, and the phase of the carrier sinusoids reverses when the sign of the error signal reverses. Therefore, demodulators used

in servomechanisms must have zero output when the input is zero, and the polarity of the demodulator d-c output must be determined by the phase of the input sinusoids, with respect to the carrier. Because of this requirement, the demodulators are called *phase-sensitive detectors*. Since the input to the demodulator contains only frequency terms which are the sum and difference of the carrier and modulating frequencies, the demodulator must be separately supplied with a signal at the carrier frequency (carrier reinsertion). Commonly used modulators are described in Par. 12-2. Paragraph 12-3 discusses demodulators.

12-2 MODULATORS

12-2.1 CHOPPER MODULATORS

Choppers are electromechanical modulators and, because of their simplicity and high zero stability (very nearly zero output when the signal to be modulated is zero), find wide application in servomechanisms.

12-2.2 Description

The essential elements of an electromechanical chopper are shown schematically in Fig. 12-2. A source of periodic voltage, usually sinusoidal, is connected to the drive coil. The resultant magnetic field causes the reed to vibrate between the upper and lower contacts at a frequency equal to that of the driv-

ing voltage. When V , the voltage to be modulated, is connected as shown in Fig. 12-3, the square-wave output waveform has the

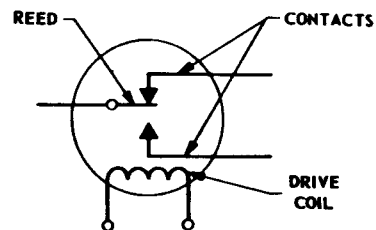


Fig. 12-2 Chopper elements.

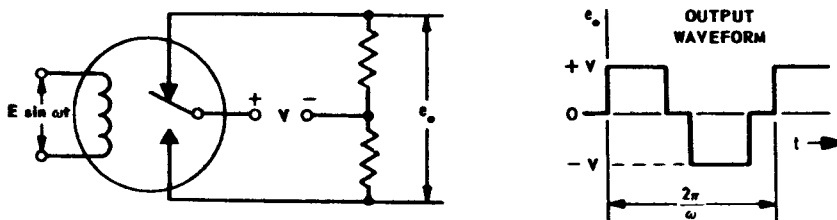


Fig. 12-3 Chopper modulator.

same period as the driving voltage, and the magnitude of each of the positive and negative rectangles of the waveform equals V . The duration of the intervals when there is zero output is a function of the time spent by the reed in moving from one contact to the other. If d denotes the fraction of each half-cycle during which the output is zero, then the output voltage e_o is given by

$$e_o(t) = \frac{4V}{\pi} \sum_{m=1}^{\infty} \frac{1}{2m-1} \cos \left[d \frac{\pi}{2} (2m-1) \right] \sin [\omega(2m-1)t] \quad (12-1)$$

where zero time reference is taken midway between the occurrence of $+V$ and $-V$ at the output. This series contains only odd harmonics of the driving-voltage frequency (fundamental, third, fifth, etc.); therefore, it is not difficult to filter the chopper output to isolate the desired fundamental-frequency component.

12-2.3 Characteristics

Figure 12-4 shows the chopper output waveform with the driving voltage superimposed. (In Fig. 12-4, the waveform of the output voltage is somewhat idealized. After initial closure, a contact usually bounces once or twice, so that the output waveform contains a few gaps at the leading edge of each rectangle.) The duration of each rectangle is called the *dwell-time*, which varies with different designs, but is typically 150° .

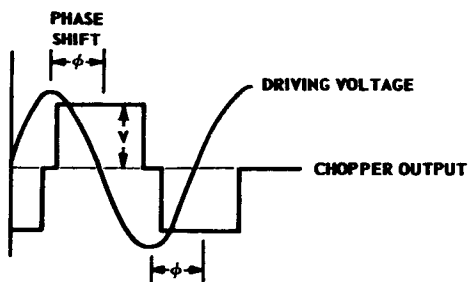


Fig. 12-4 Relationship between driving voltage and chopper output.

12-2.4 Phase shift ϕ . Of particular interest to the user is the phase shift ϕ between the driving voltage and the modulator output. An example is the application in which the chopper is used to modulate a d-c error signal which is then amplified and used to drive a 2-phase motor. The motor responds only to signals that are 90° out of phase with the motor reference-winding voltage, which frequently is also used to drive the chopper. Hence, the over-all phase shift of the chopper and the amplifier must be adjusted to 90° . For this reason, the phase-shift characteristics of the chopper must be known.

Chopper phase shift ϕ is defined as the number of degrees of the driving voltage by which the mid-point between contact make-and-break lags the peak of the driving voltage. The lag is due to the inductive nature of the driving coil and also to the mechanical constants of the reed. The lag angle may be adjusted by inserting a phase-shifting device such as a capacitor, resistor, or a parallel combination of both, in the drive-coil circuit. When these components are used, the input-voltage amplitude must be adjusted so that the rated coil voltage is applied to the coil. The phase angle is a function of temperature, drive-voltage amplitude, and drive-voltage frequency. Where these parameters are expected to vary in an application, the chopper manufacturer should be consulted for quantitative information about the effects of these variations. The phase angle can be substantially stabilized against changes in the drive-voltage frequency by the addition of a resistor in series with the drive coil.

12-2.5 Drive-voltage frequency. Commercially available electromechanical choppers are designed for drive-voltage frequencies ranging up to several thousand cps. However, most units are designed for 60- or 400-cps operation.

12-2.6 Drive voltage. The most commonly specified drive-coil voltage is 6.3 volts, but models with somewhat higher ratings are also made. Typical drive-coil power requirements range from 0.5 to 1 watt.

12-2.7 Contact rating. The useful life of a chopper is affected most by the contact rating. Most chopper failures are caused by contact wear, pitting, and sticking or the development of contact resistance. These failures can be reduced by strict adherence to the maximum contact ratings which, in typical commercial units, are a few volts and a few milliamperes. Since useful life depends upon how far chopper characteristics can deteriorate before the servomechanism fails, the number of hours of satisfactory operation will vary with the application. It is generally true, however, that where rated-performance limits are not exceeded, several thousand hours of life can be expected.

12-2.8 Chopper noise. Noise generated by an electromechanical chopper varies with the design of the unit, the impedance of the circuit connected to the output, and the bandwidth of the measuring equipment. Since the bandwidth of a servomechanism is small, chopper noise is usually not troublesome, and can be kept low by connecting the output of the chopper to a circuit of moderately low impedance. One design, for example, has a rated noise level of 50 microvolts peak-to-peak when connected to a 2200-ohm load, but this level jumps to 1.5 millivolts when connected to a 1-megohm load. These figures are for a bandwidth of 0.2 to 1000 cps, which is wider than that of a servomechanism.

12-2.9 Temperature effects. Choppers are rugged devices that will operate satisfactorily over a considerable range of temperatures. While temperature does affect phase lag, choppers will operate satisfactorily over a temperature range which, for some designs, is as wide as -55° to $+200^{\circ}\text{C}$. Hermetically sealed models are available which are rated for operation at altitudes as high as 50,000 feet and in a high-humidity environment. The ruggedness of choppers is illustrated by the vibration rating of one design, which is 30 g's for frequencies between 5 and 500 cps.

12-2.10 Packaging

Most choppers are housed in cylindrical

cans, one to two inches in diameter, and one to four inches long.

12-2.11 Practical Circuits

Figures 12-5 through 12-8 show practical connection circuits of chopper contacts, and the corresponding output waveforms. All waveforms have a period $1/f$, where f is the frequency of the drive-coil excitation voltage. The circuit of Fig. 12-5 is more suitable than that of Fig. 12-6 when the source impedance of V is high. The circuit of Fig. 12-7 produces a push-pull output and is therefore useful when a double-ended signal is desired. Figure 12-8 shows a circuit in which the peak-to-peak output voltage is the difference between the two input signals. This circuit may therefore be used to modulate the error signal in a servomechanism when e_{IN1} represents the command and e_{IN2} represents the response.

12-2.12 MAGNETIC MODULATORS^(1, 2)

A magnetic modulator converts the d-c signal to a modulated a-c signal with a carrier frequency twice that of the modulator excitation voltage. The output of the magnetic modulator is phase-sensitive to the polarity of the d-c signal, permitting demodulation of the amplified signal by a phase-sensitive detector and restoration of the original polarity at the output of the amplifier.

12-2.13 Principle of Operation

The basic circuit of a magnetic modulator is shown in Fig. 12-9. The circuit consists of two wound ferromagnetic, high-permeability cores, which are closely matched for magnetic characteristics, each carrying a signal winding, an excitation winding, and an output, or load, winding. The load windings are connected in opposite polarity sense with respect to the excitation windings so that, with no signal current, the load-winding voltages e_1 and e_2 are identical but opposite in phase, thereby cancelling to produce zero load voltage. When a control current is applied from the signal source, the control current produces asymmetry in the operation of the two

SIGNAL CONVERTERS

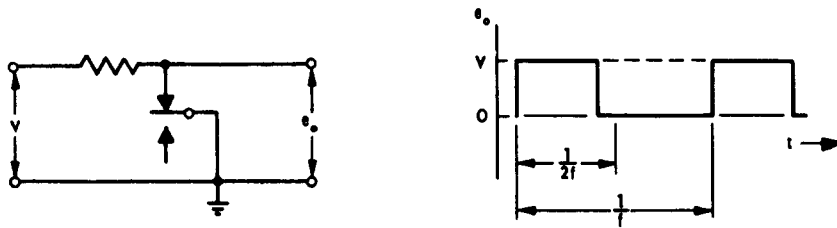


Fig. 12-5 Half-wave connection.

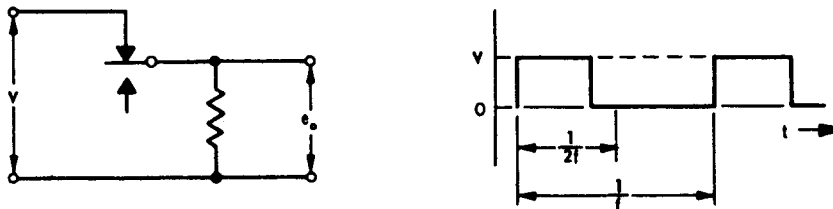


Fig. 12-6 Half-wave connection.

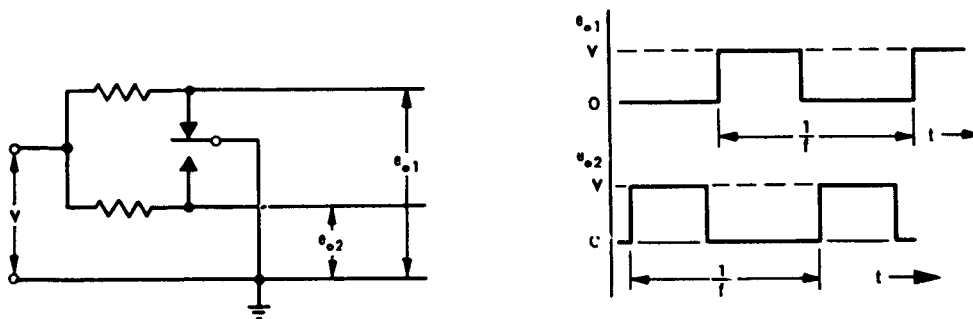


Fig. 12-7 Push-pull half-wave connection.

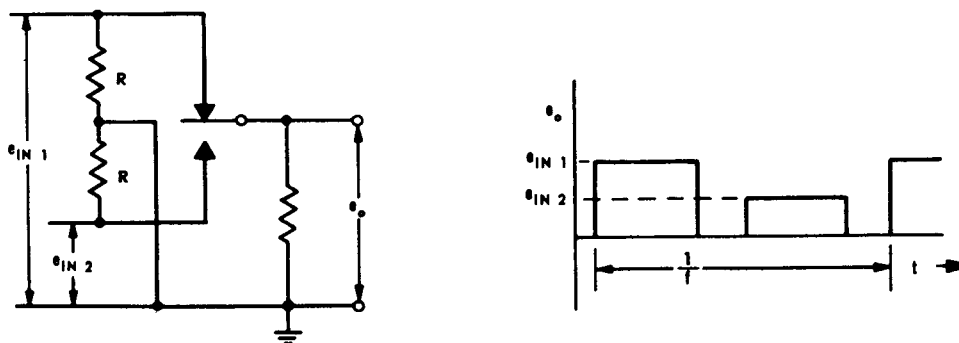


Fig. 12-8 Differential connection.

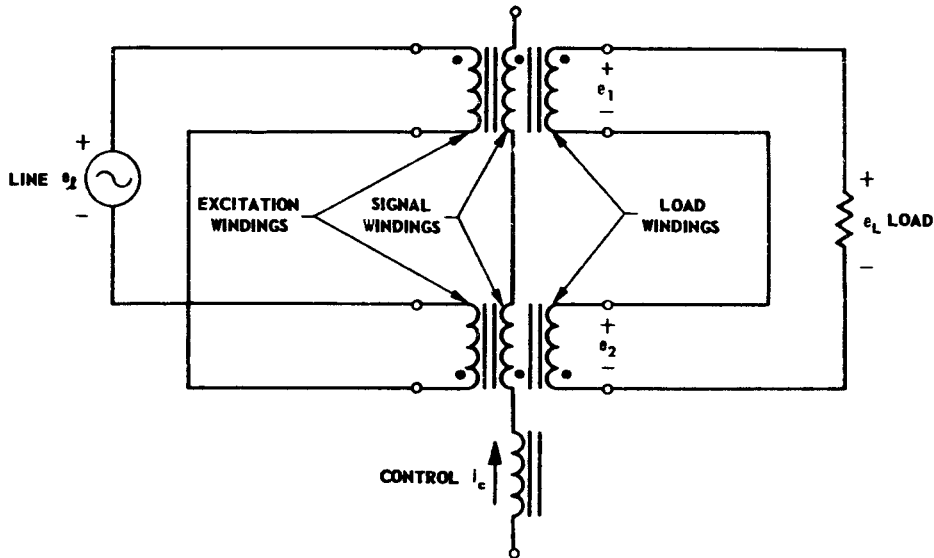


Fig. 12-9 Basic circuit of second-harmonic modulator.

cores, and a voltage is produced in the load winding at twice the frequency of the excitation source and proportional to the magnitude of the control signal.

An understanding of the manner in which this second-harmonic output is produced can be gained by referring to the hysteresis loop of the core material (see Fig. 12-10). The signal current can be assumed to produce a displacement of the initial flux operating point of the core material, from point 1 to the points on the hysteresis loops for the corresponding cores. As a cycle starts, core 1 is operating in a region of low permeability, whereas core 2 is operating in a region of high permeability. Hence, the applied line voltage e_1 appears principally across core 2, and the load-winding voltage e_2 (as shown in Fig. 12-11) is larger in magnitude than load-winding voltage e_1 at that time in the cycle. As the cycle progresses, and the operating points of the core material move toward positive saturation, core 2 encounters a region of low permeability near the knee of the saturation curve, while core 1 is now in a region of high permeability. Voltages e_1 and e_2 are now reversed in relative amplitude. When the sum of voltages e_1 and e_2 is taken as a load voltage

e_L (as shown in Fig. 12-11), the net result is a second-harmonic component. If the direction of the signal current is reversed, then the cores interchange positions on the hysteresis loops, and the second-harmonic load voltage reverses in phase. This feature of the magnetic modulator makes it possible to restore the polarity of the original d-c signal by means of a phase-sensitive detector.

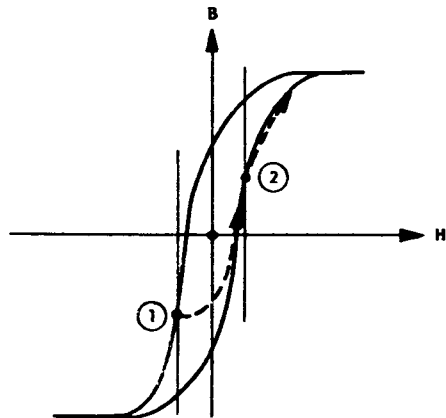


Fig. 12-10 Operation of cores for positive control signal.

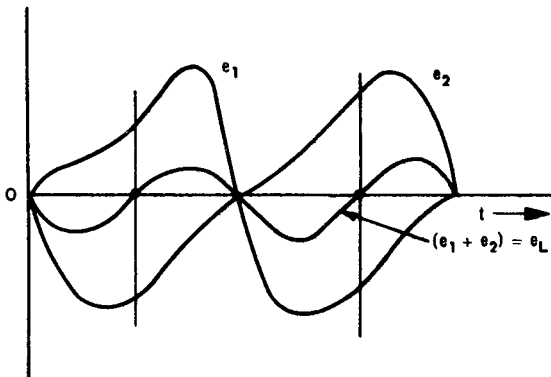


Fig. 12-11 Secondary voltage for $i_c = I_c$ condition.

12-2.14 Operating Circuits

There are two categories of operating circuits employing the magnetic modulator: one category is very similar to that shown in Fig. 12-9, in which the signal and load windings are separated; in the second category, the d-c signal and the load are connected to a common winding, each core having only two windings — an excitation winding and a signal-load winding.

12-2.15 Sensitivity

For maximum magnetic-modulator sensitivity, the control-winding impedance is usually matched to the impedance of the signal source, and the load-winding impedance is usually matched to the load impedance, although this may not be true with high-impedance load circuits, such as grid circuits of vacuum-tube amplifiers. It is possible to secure individual load-impedance and signal-impedance matching with the circuit in which the control and load windings are separated. However, the common-winding circuit will result in greater sensitivity and, thus, may be more desirable in certain applications.

12-2.16 Harmonic Attenuation

One feature found in almost all magnetic modulators is a capacitor for tuning the load to the second-harmonic frequency, so that odd harmonics and extraneous noise are attenuated.

12-2.17 Performance and Application of Magnetic Modulators

The control characteristics of a typical, commercially available magnetic modulator are shown in Fig. 12-12. This modulator operates with an excitation voltage of 5 volts in the 300- to 500-cps frequency range. The weight of the unit is approximately 3 ounces. The range of linear output, approximately 5 volts rms, is obtained with about 100 microamperes of signal into a resistance of 1000 ohms, which is equivalent to a signal power of 10^{-5} watt. The limitation on minimum signal power required to operate a magnetic modulator is set by the inherent Barkhausen noise, which is in the order of 10^{-19} watt per cycle of bandwidth. Some workers⁽³⁾ have reported the design of magnetic modulators in which the zero error can be reduced to 5×10^{-17} watt input, with a random variation of 3×10^{-18} watt over a two-hour period. Two possible hindrances to building extremely sensitive magnetic modulators are: the harmonic level of the oscillator that supplies the excitation power; and the mismatch, or magnetic asymmetry, in the two cores.

12-2.18 Core Material

Raising the excitation frequency results in higher gain and a reduced time constant for the magnetic modulator, but it may also result in excessive core loss. However, the excitation voltage should saturate the core material to realize optimum performance. Therefore, a suitable core material must have a high maximum permeability and low core loss. A good core material that is not too susceptible to mechanical shock is 4-79 Mo-Permalloy.

12-2.19 ELECTRONIC MODULATORS⁽⁴⁾

Circuit schematics of electronic modulators commonly used in servomechanisms are shown in Figs. 12-13 through 12-16.

12-2.20 Operation

In all of these circuits, except Fig. 12-16, the principle of operation is the same and is similar to chopper operation. Nonlinear elements, either diodes or triodes, are used as switches to connect the d-c signal V to the

MEASUREMENT AND SIGNAL CONVERTERS

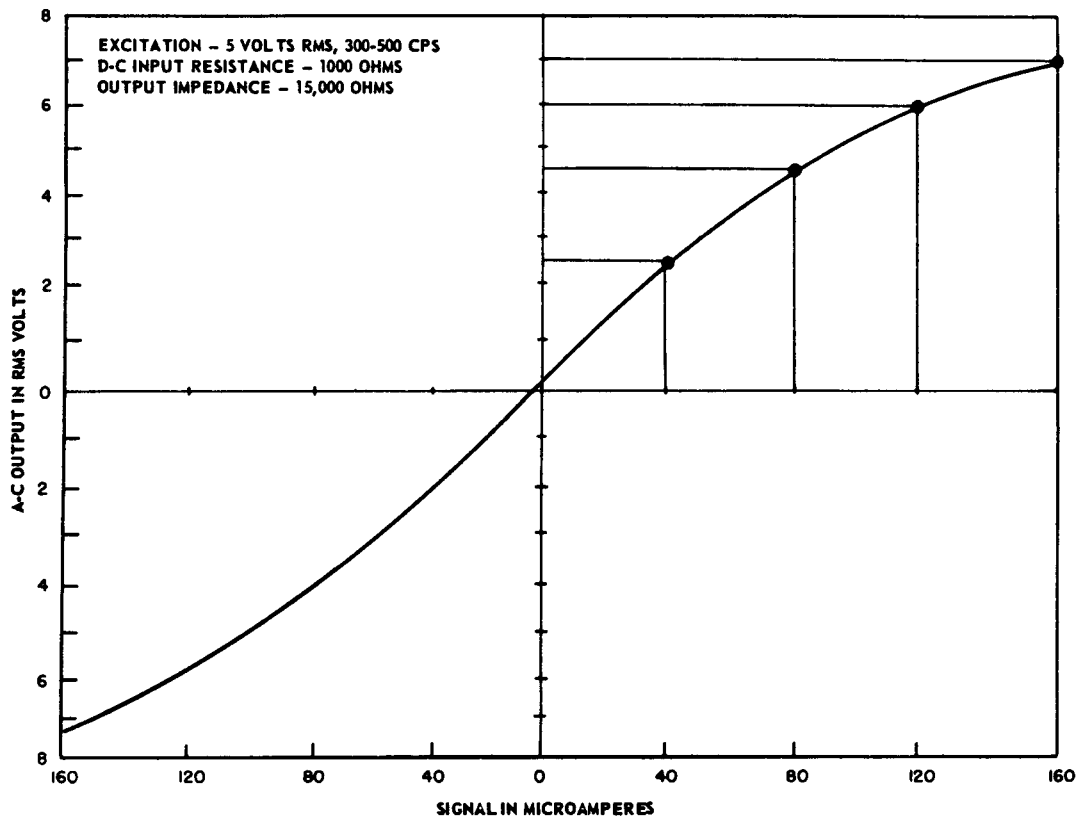


Fig. 12-12 Control characteristics of commercial magnetic modulator.

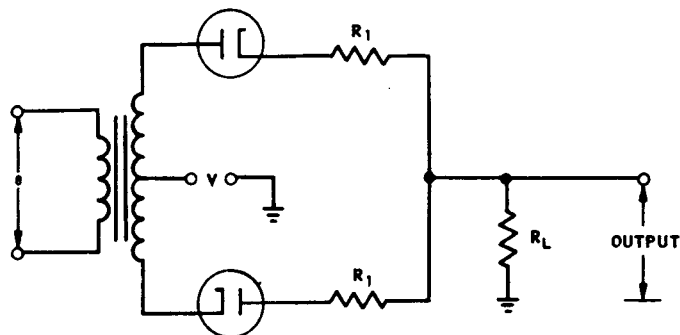


Fig. 12-13 Diode modulator.

SIGNAL CONVERTERS

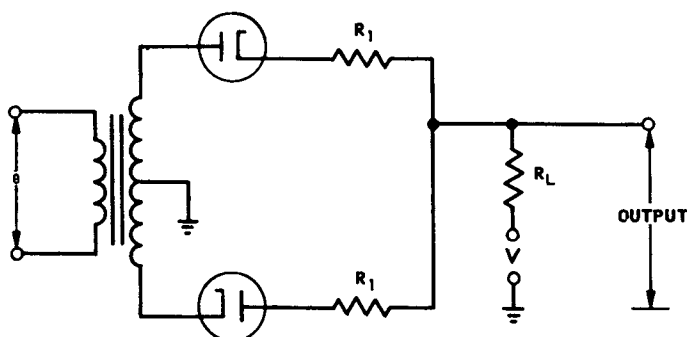


Fig. 12-14 Diode modulator.

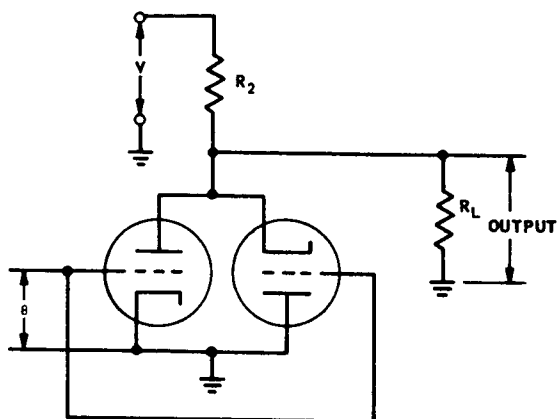


Fig. 12-15 Triode modulator.

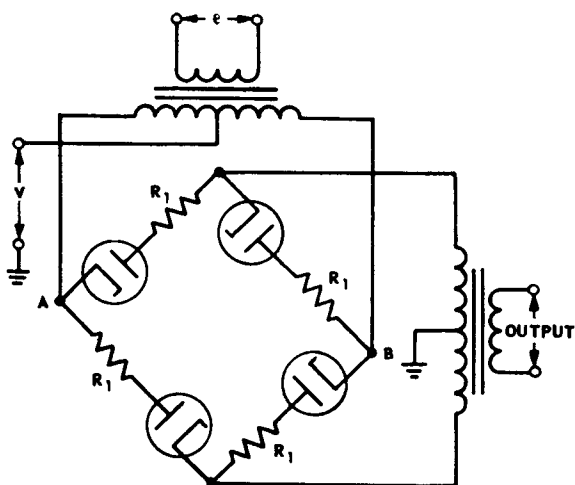


Fig. 12-16 Bridge modulator.

output during one half-cycle of the sinusoidal carrier voltage e . During the other half-cycle of e , the output is connected to ground. Therefore, the output waveform is a rectangle of amplitude V , with a duration of half a period of the carrier and with a repetition rate equal to the modulating frequency. Resistors R_1 are connected in series with all diodes to limit the diode current when the

diodes conduct. The value of R_1 should be much less than that of R_L , a typical value being 1000 ohms.

In Fig. 12-13, the rectangle occurs (i.e., the diodes conduct and connect V to the output) when the upper end of the transformer secondary is positive. During the other half-cycle, the output is at ground potential. In Fig. 12-14, the output is connected to ground

when the upper end of the transformer secondary is positive, and the rectangle occurs during the other half-cycle of the carrier. In a variation of this circuit, a second d-c signal is connected to the center tap of the transformer. The output signal then has a peak-to-peak amplitude that is the difference between the two d-c signals, and the modified circuit can therefore be used as a combination error (difference) detector and modulator. In the back-to-back triode circuit of Fig. 12-15, a very large carrier signal is applied to the grids, making one of the two triodes conduct during approximately the entire positive half-cycle of the carrier signal. The left tube conducts during the positive half-cycle if V is positive and the right tube conducts if V is negative. The grid signal is so large that the tube drop is very small during conduction and, therefore, the output is nearly at ground potential during conduction. In order for this to occur, R_2 and R_L must be large (e.g., 500,000 ohms). The circuit is used only when large signals, in the order of 100 volts, are to be modulated.

12-2.21 Bridge modulator. Figure 12-16 differs from the other configurations in that two rectangles of opposite polarity are generated during each cycle of the carrier. When the carrier polarity is such that point A is positive with respect to point B, V appears across the output transformer with one polarity; if A is negative with respect to B, V appears across the output transformer with opposite polarity.

12-2.22 Typical curve. In the diode circuits, the statement that the rectangles have a duration of half a period of the modulating signal is based on the assumption that the peak amplitude of e is large compared with V . Where e is small compared with V , saturation occurs. A typical curve is given in Fig. 12-17, showing the rms value of the output voltage as a function of the magnitude of V for a fixed amplitude of e . The knee in the curve shows the occurrence of saturation,

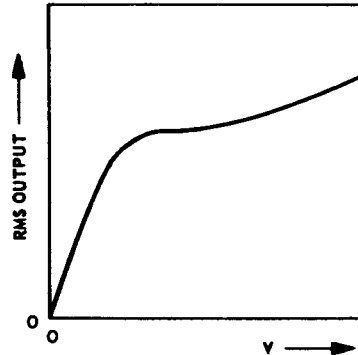


Fig. 12-17 Modulator rms output as a function of amplitude of V for fixed amplitude of e .

which must be avoided if the output is to be proportional to V .

12-2.23 Stability. In all of the diode circuits shown, zero stability (zero input voltage) depends upon the balance between the diodes. For example, in Fig. 12-13, assuming that V is zero, the output will be at zero potential during the half-cycle of the carrier when the upper end of the transformer secondary is positive, but only if the drop is equal in both diodes and their associated resistors R_1 . Therefore, zero stability is determined by how nearly identical are the two diodes. For good zero stability, the diodes must be identical, not only when the circuit is first adjusted, but also throughout the aging process of the diodes. The diodes need not be vacuum tubes; they can be semiconductor devices, such as germanium or silicon diodes. When semiconductor diodes are used, it must be remembered that they have measurable reverse current. It is important that the diodes be matched in their forward as well as reverse characteristics. In general, zero stability of vacuum-tube modulators is not as good as that of chopper modulators. Therefore, chopper modulators are preferred for high-precision applications.

12-3 ELECTRONIC DEMODULATORS^(5,6)

12-3.1 DIODE DEMODULATORS

Figure 12-18 shows the schematic of a diode circuit frequently used in servomechanisms for demodulating suppressed-carrier signals; i.e., signals in the form

$$e_{IN} = A \sin \omega_s t \sin \omega_c t \quad (12-2)$$

where

ω_s = angular frequency of modulated signal (e.g., error signal)

ω_c = angular frequency of carrier

This signal is connected to the primary of the center-tapped transformer. A signal corresponding to the carrier

$$e_c = B \sin \omega_c t \quad (12-3)$$

is connected to the other transformer. The secondary voltages corresponding to e_{IN} and e_c are, respectively

$$e_1 = V_1 \sin \omega_s t \sin \omega_c t \quad (12-4)$$

$$e_2 = V_2 \sin \omega_c t \quad (12-5)$$

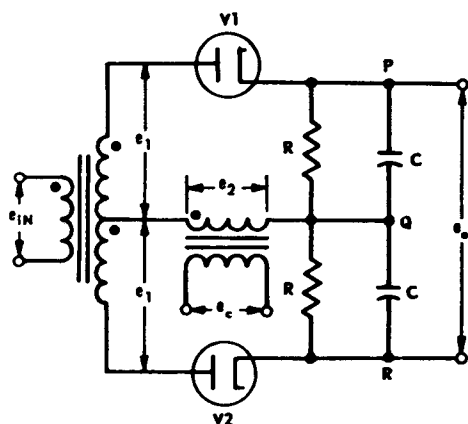


Fig. 12-18 Diode demodulator.

12-3.2 Operation

The operation of the diode demodulator (Fig. 12-18) is as follows: The voltage between Q and the plate of diode $V1$ is the sum of e_2 and e_1 . The voltage between Q and the plate of diode $V2$ is the difference between e_2 and e_1 . Hence, when e_1 is zero, the voltage across PQ equals the voltage across RQ , and the output voltage e_o (which is the difference between the two voltages) is zero. If e_1 is in phase with e_2 ($\sin \omega_s t$ is positive), the voltage across PQ is greater than the voltage across RQ , and the output voltage e_o is positive. If e_1 is 180° out of phase with e_2 ($\sin \omega_s t$ is negative), the voltage across PQ is smaller than the voltage across RQ , and the output voltage e_o is negative. In general, ignoring ripple in the output due to discharging of capacitors C between conduction periods of the diodes, and ignoring the time lag between input and output, it follows that

$$e_o = 2V_1 \sin \omega_s t \quad (12-6)$$

provided V_1 equals or is less than V_2 . For V_1 greater than V_2 , saturation occurs. If e_1 contains a quadrature component (a component that is phase-shifted 90° with respect to the carrier), the output due to the quadrature component is zero.

12-3.3 Ripple. The output voltage contains a ripple component (the waveform of which approximates a saw tooth) that is amplitude modulated by the variations in the modulated signal. The ripple is caused by the discharge of capacitors C between peaks of the carrier signal. The capacitors are charged during peaks, when the diodes conduct. The output ripple decreases when the RC product increases but, as this product is increased, the output voltage is no longer able to follow variations in the modulated signal. An upper limit for RC is given by

$$RC \leq \frac{1}{m\omega_s} \sqrt{1 - m^2} \quad (12-7)$$

where

$$m = \frac{V_1}{V_2}$$

The amplitude of the first harmonic of the ripple, which has the same frequency as the carrier, is given by

$$N_1 = \frac{2V_1}{f_c} \sqrt{\frac{1}{(\pi RC)^2} + (2f_s)^2} \quad (12-8)$$

where

$$f_c = \frac{\omega_c}{2\pi}$$

$$f_s = \frac{\omega_s}{2\pi}$$

The ripple factor is defined as the ratio of the amplitude of the first harmonic of the ripple to the amplitude of the output fundamental and, for values not in excess of 10 percent, is given by

$$r = \frac{1}{f_c} \sqrt{\frac{1}{(\pi RC)^2} + (2f_s)^2} \quad (12-9)$$

Since f_s and f_c are fixed in a given application, the specification for r determines the required RC product. The inequality [Eq. (12-7)] then determines the maximum value of m and, hence, the upper limit on the ratio of V_1 to V_2 is determined for the design. As r decreases, the RC product increases, and the value of m decreases.

12-3.4 Unbalance. The parallel impedance of R and C at ω_c should be large compared with the equivalent source impedances e_1 and e_2 and compared with the forward resistance of the diodes so that transformer and diode unbalance is of no practical significance.

12-3.5 Transfer function. An expression for the output voltage that is more accurate than the one stated previously is given by the following transfer function for the demodulator:

$$\frac{E_o(s)}{2E_1(s)} = \frac{1 - \frac{1}{2f_c RC}}{1 + T_d s} \quad (12-10)$$

where

$$T_d = \frac{\frac{1}{2f_c}}{1 - \frac{1}{2f_c RC}} = \text{demodulator time constant}$$

s = complex frequency of modulated signal

12-3.6 Output stability. The demodulator output should always be taken across P and R , as shown in Fig. 12-18. If push-pull outputs are taken across PQ and RQ separately, the ripple at zero input signal results in time-varying outputs and poor stability.

12-3.7 Full-Wave Demodulator

Figure 12-19 shows the schematic diagram of a full-wave diode demodulator. In this circuit, which uses four diodes, conduction occurs during both halves of the carrier cycle. The demodulator time constant and ripple factor are reduced by a factor of 2 over the half-wave circuit of Fig. 12-18. An equivalent full-wave diode demodulator is shown in Fig. 12-20.

12-3.8 TRIODE DEMODULATORS

The diodes in Fig. 12-18 may be replaced by triodes. This produces the circuit of Fig. 12-21, which is characterized by amplification of the modulated signal and by high input impedance for e_{ix} . Quadrature voltages, however, are not rejected; for this reason, the circuit is rarely used in servomechanisms.

12-3.9 Keyed Demodulator

A different demodulator configuration, called a *keyed* demodulator, is shown in Fig. 12-22. It is characterized by smaller ripple for a given speed of response.

12-3.10 Operation. The operation of a keyed demodulator is as follows: The transformer supplies large in-phase signals, corresponding to the carrier, to both grids. Self-bias is developed through grid conduction in C_g , and the time constant $R_g C_g$ is chosen to permit very little bias loss between carrier cycles. As a result, the tubes can conduct only during the positive peak of the carrier voltage. At that time, the input is connected to the output, and the demodulator is said to be

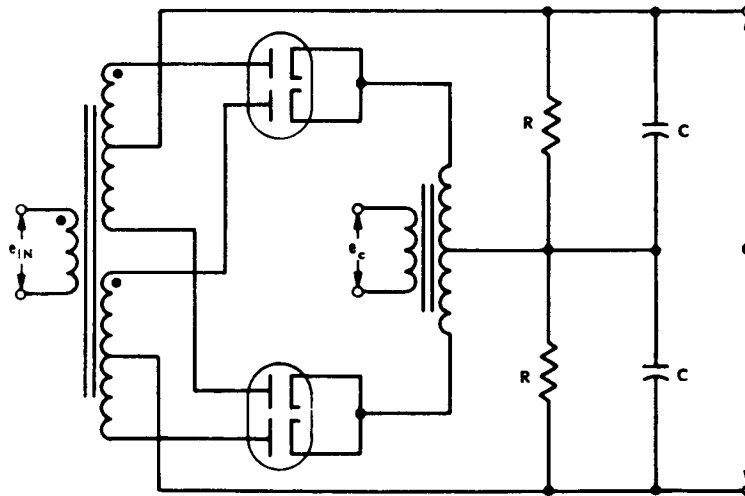


Fig. 12-19 Full-wave diode demodulator.

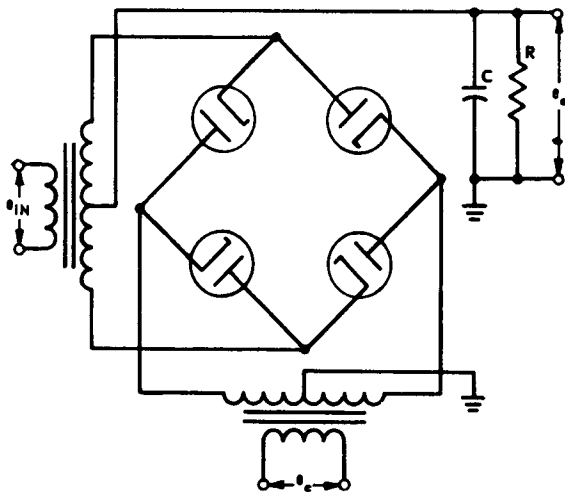


Fig. 12-20 Full-wave (ring) diode demodulator.

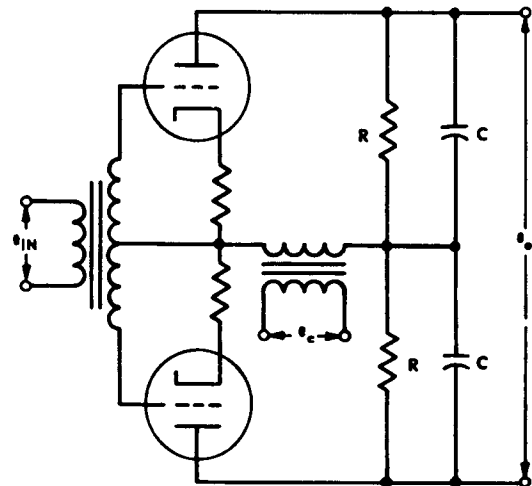


Fig. 12-21 Triode demodulator.

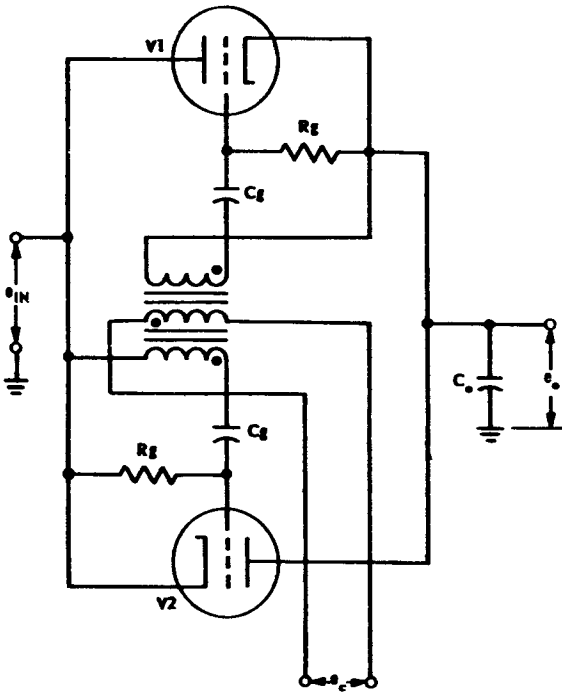


Fig. 12-22 Keyed demodulator.

keyed by the peak of the carrier signal. Tube V1 conducts during keying if the input is *larger* than the output, and V2 conducts if the input is *smaller* than the output. The output capacitor C_o charges to the peak input voltage, which is the instantaneous value of the modulated signal and, therefore, the output voltage is a replica of the modulated signal. If the modulated signal is a sinusoid, the output is also a sinusoid, but with a lag of $1/2f_c$ seconds and including a ripple of saw-tooth shape.

12-3.11 Transfer function. The transfer function for the keyed demodulator is

$$\frac{E_o(s)}{E_{IN}(s)} = \frac{1}{1 + T_d s} \quad (12-11)$$

where

$$T_d = \frac{1}{2f_c} = \text{demodulator time constant}$$

The ripple voltage is zero when the modulated signal is constant. When the modulated signal is a sinusoid, the ripple voltage is a saw tooth, amplitude modulated by that sinusoid. The ripple factor, which is lower than that of the diode demodulator, is

$$r = \frac{2f_s}{f_c} \quad (12-12)$$

Quadrature rejection is equally good, however, for both the keyed demodulator and the diode demodulator. The triodes of the keyed demodulator can be replaced by transistors or magnetic amplifiers.⁽⁷⁾

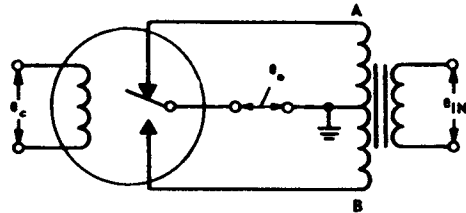


Fig. 12-23 Full-wave chopper demodulator.

12-3.12 CHOPPER DEMODULATORS

Choppers can also be used as demodulators. Figure 12-23 shows the schematic diagram of a full-wave chopper demodulator.

12-3.13 Operation

The modulated signal e_{IN} is connected to the primary of the center-tapped transformer and the carrier voltage e_c is connected to the drive coil of the chopper. The chopper connects the output to terminal A of the transformer during the first half-cycle of the carrier and to terminal B during the other half-cycle. Thus, if e_{IN} is phased with respect to the reed motion so that A is positive during the first half-cycle, B will also be positive during the second half-cycle, and the output is a positive voltage. If A is negative during the first half-cycle, B will also be negative during the second half-cycle, and the output is a negative voltage. The polarity of the output voltage e_o is, therefore, a function of the

SIGNAL CONVERTERS

phase relationship between the reed motion and the signal to be demodulated. If these two are 90° out of phase, the output is zero. If they are either in phase or 180° out of phase, the magnitude of the output voltage is determined by the magnitude of the modulated signal.

12-3.14 Output filtering. The output voltage contains a ripple voltage that varies at a frequency f_c and also at harmonics of f_c . Therefore, a filter is usually required at the output of the demodulator. The design of this filter becomes increasingly less difficult as the highest-frequency component of the modulated

signal is made smaller and smaller compared with the carrier frequency. In typical applications, ω_s is less than 10 percent of ω_c , and the filter design is not difficult.

12-3.15 Dual Use

When a chopper is used as a modulator, the same unit can also be used as a demodulator. This is shown in Fig. 12-24. The reed and contact *A* of the chopper form a half-wave modulator, like the one shown in Fig. 12-5. The reed and contact *B* of the chopper form a half-wave demodulator, which permits the output of the amplifier to pass to

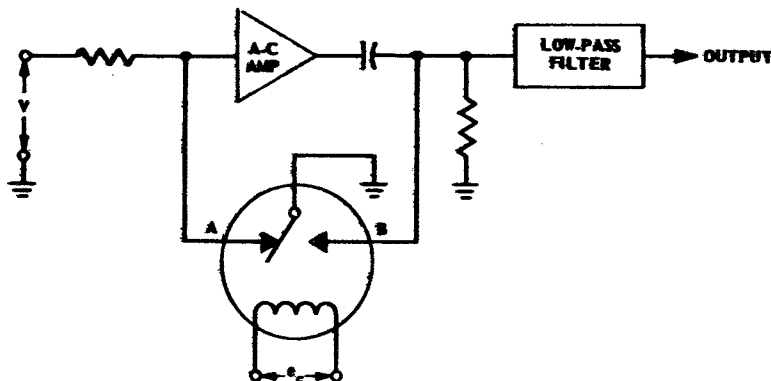


Fig. 12-24 Single chopper used as both modulator and demodulator.

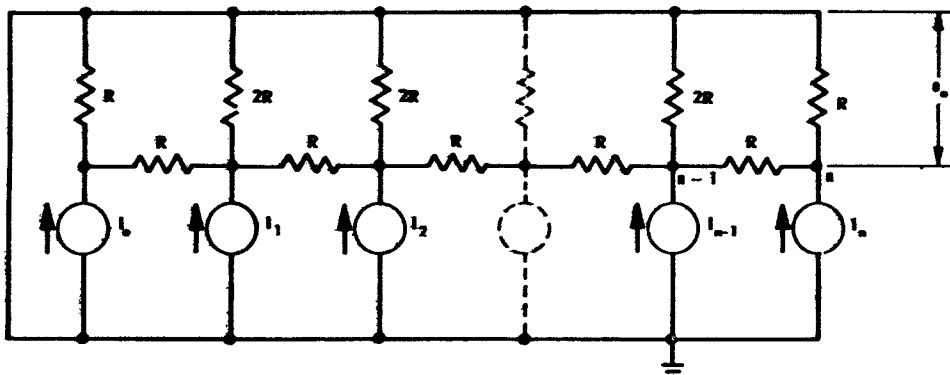


Fig. 12-25 Digital-to-voltage converter.

Adapted by permission from *Notes on Analog-Digital Conversion Techniques*, edited by A. K. Susskind, 1957, Technology Press, Massachusetts Institute of Technology.

the filter during one-half-cycle of reed motion and grounds the filter input on the other half-cycle. This results in a unidirectional filter-input signal, the magnitude of

which is determined by the magnitude of the amplifier output, and the polarity of which is determined by the phase of the amplifier output with respect to the reed motion.

12-4 DIGITAL-TO-ANALOG CONVERSION

12-4.1 GENERAL

Control systems are now coming into use in which digital data-processing equipment is employed. In digital equipment, information is expressed as numbers. In the rest of the system, information is represented by analog quantities, such as voltage or shaft rotation. In order to couple the output of the digital equipment to the other parts of the system, digital-to-analog conversion must be performed.

12-4.2 Networks

Where the required analog representation is to be a voltage, and accuracy requirements are not very high, electrical networks can be used. Examples of several network configurations are shown in Figs. 12-25 through 12-28.

12-4.3 Operation. In Fig. 12-25, each current source I_i has an output current I when the associated binary digit (value 2^i) is a ONE, but otherwise has zero output. If the current sources are assumed to have infinite impedance, the output voltage is given by

$$e_o = \frac{IR}{3 \times 2^{n-1}} p \quad (12-13)$$

and the output impedance is given by

$$R_o = \frac{2}{3} R \quad (12-14)$$

where

p = number to be converted

$n + 1$ = number of stages

In Fig. 12-26, each voltage source E_i has an output voltage E when the associated binary digit is a ONE, but otherwise has zero output. If the voltage sources are assumed to have zero impedance, the output voltage is given by

$$e_o = \frac{E}{2^{n+1}} p \quad (12-15)$$

and the output impedance is given by

$$R_o = \frac{R}{2} \quad (12-16)$$

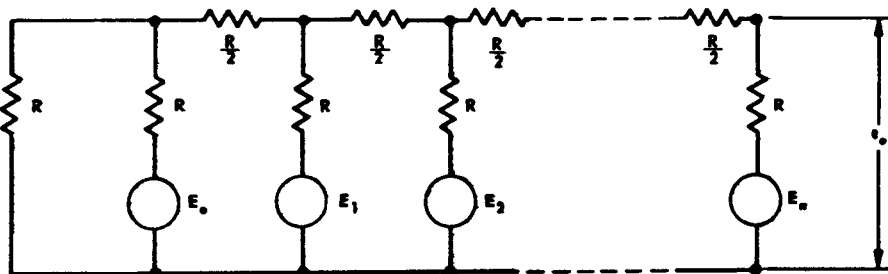


Fig. 12-26 Digital-to-voltage converter.

SIGNAL CONVERTERS

Figure 12-27 shows a circuit that is particularly useful when the digital information originates from relays. Here, relay K_i is not energized (contacts in the position shown) when the associated binary digit is a ZERO. The contacts are in the other position, however, when the associated digit is a ONE. The output voltage is given by

$$e_o = \frac{E_1}{\frac{R}{R_c} + 2^{n+1} - 1} p \quad (12-17)$$

and the output impedance is given by

$$R_o = \frac{R}{2^{n+1} - 1} \quad (12-18)$$

Figure 12-28 shows another circuit that is also useful when the digital information originates from relays. Here again, relay K_i is not energized (contacts in the position shown) when the associated binary digit is a ZERO, and relay K_i is energized (contacts in the other position) when the associated binary digit is a ONE. The output voltage is given by

$$e_o = \frac{E}{2^{n+1} - 1} p \quad (12-19)$$

12-4.4 Accuracy. The circuits of Figs. 12-27 and 12-28 have the advantage over those of Figs. 12-25 and 12-26 because they require only one precision resistor per stage. In all of the circuits, accuracy is determined by the

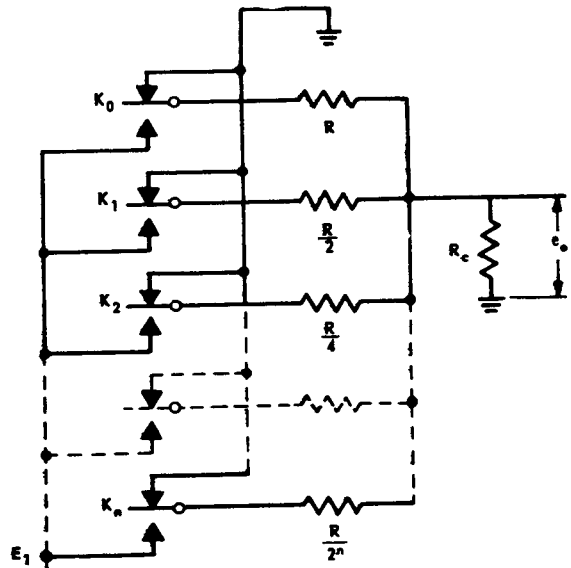


Fig. 12-27 Digital-to-voltage converter.

Adapted by permission from *Notes on Analog-Digital Conversion Techniques*, edited by A. K. Susskind, 1957, Technology Press, Massachusetts Institute of Technology.

accuracy of the precision resistors. In addition, in Figs. 12-25 and 12-26, accuracy is affected by the magnitude of the outputs of the sources and also by their output impedance. The circuits of Figs. 12-27 and 12-28 are therefore somewhat more accurate, but even these can rarely be made more accurate than

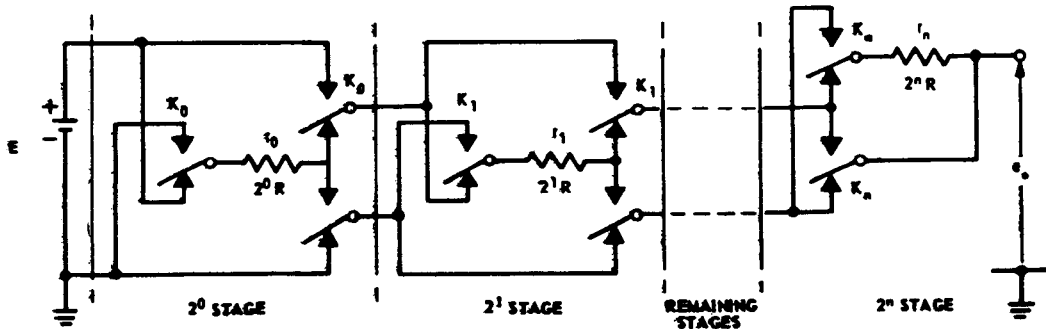


Fig. 12-28 Digital-to-voltage converter.

Adapted by permission from *Notes on Analog-Digital Conversion Techniques*, edited by A. K. Susskind, 1957, Technology Press, Massachusetts Institute of Technology.

one part in 4000 (12 binary digits). For the circuits of Figs. 12-25 and 12-26, typical accuracy figures are one part in 1000 (10 binary digits).

12-4.5 Servomechanism

Where higher-accuracy requirements must be met, a servomechanism can be used. A block diagram of a digital-to-analog servomechanism is shown in Fig. 12-29. Here, the position of the output shaft is expressed in digital form by a coding disc, such as the one

described in Ch. 11. The coded shaft position is then subtracted from the desired shaft position, and the difference number is converted into a voltage which, after dynamic compensation and amplification, is used to drive the output motor. The accuracy of conversion is determined by the coding-disc sensing element. Since coding discs can be made with an accuracy as high as one part in 65,536 (16 binary digits), it follows that the static conversion accuracy in the closed-loop approach can be made very high.

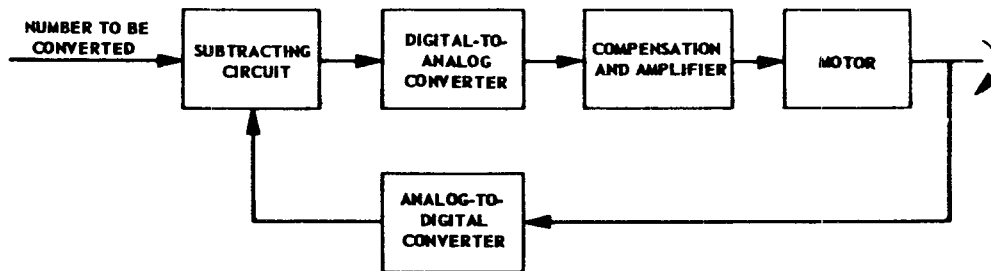


Fig. 12-29 Servomechanism for digital-to-analog conversion.

BIBLIOGRAPHY

- 1 W. A. Geyger, *Magnetic Amplifier Circuits*, Ch. 16, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.
- 2 G. Wennerberg, "A Simple Magnetic Modulator for Conversion of Millivolt Direct Current Signals", *Electrical Engineering*, Vol. 70, No. 2, pp. #144-147, February, 1951.
- 3 S. C. Williams and S. W. Noble, "The Fundamental Limitations of the Second Harmonic Type of Magnetic Modulator as Applied to the Amplification of Small Direct Current Signals", *Proc. Inst. Elec. Engrs. (London)*, Vol. 97, Part 2, No. 58, pp. #445-459, August, 1950.
- 4 W. R. Ahrendt, *Servomechanism Practice*, Ch. 5, McGraw-Hill Book Company, Inc., New York, N. Y., 1954.
- 5 K. E. Schreiner, "High-Performance Demodulators for Servomechanisms", *Proc. National Electronics Conference*, Vol. 2, 1946.
- 6 S. P. Detwiler, "Phase-Sensitive-Detector Characteristics", *AIEE Miscellaneous Paper* 51-349.
- 7 R. M. Mark, W. X. Johnson, and P. R. Johannesen, "Magnetically Keyed, Phase Sensitive Demodulators", *AIEE Conference Paper* CP 56-757.

AMC PAMPHLET

AMCP 706-138

THIS IS A REPRINT WITHOUT CHANGE OF ORDP 20-138, REDESIGNATED AMCP 706-138

ENGINEERING DESIGN HANDBOOK

SERVOMECHANISMS

SECTION 3, AMPLIFICATION



HEADQUARTERS
UNITED STATES ARMY MATERIEL COMMAND
WASHINGTON, D.C. 20315

30 April 1965

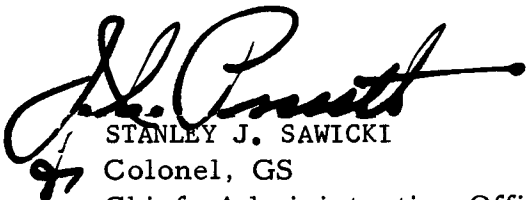
AMCP 706-138, Servomechanisms, Section 3, Amplification, forming part of the Army Materiel Command Engineering Design Handbook Series, is published for the information and guidance of all concerned.

(AMCRD)

FOR THE COMMANDER:

SELWYN D. SMITH, JR.
Major General, USA
Chief of Staff

OFFICIAL:

A handwritten signature in black ink, appearing to read "S. J. Sawicki", with a stylized flourish at the end. The signature is written over the printed name and title of Stanley J. Sawicki.

STANLEY J. SAWICKI
Colonel, GS
Chief, Administrative Office

DISTRIBUTION: Special

PREFACE

The Engineering Design Handbook Series of the Army Materiel Command is a coordinated series of handbooks containing basic information and fundamental data useful in the design and development of Army materiel and systems. The handbooks are authoritative reference books of practical information and quantitative facts helpful in the design and development of Army materiel so that it will meet the tactical and the technical needs of the Armed Forces. The present handbook is one of a series on Servomechanisms.

Section 3 of the handbook contains Chapter 13, which covers the different amplifiers used in servo controllers. This chapter describes the following types of amplifiers: electronic, transistor, magnetic, rotary electric, relay, hydraulic, pneumatic and mechanical. The advantages and disadvantages of the various types are discussed to help the circuit designer in making a choice for his particular application.

For information on other servomechanism components and on feedback control theory and system design, see one of the following applicable sections of this handbook:

AMCP 706-136 Section 1 Theory (Chapters 1-10)

AMCP 706-137 Section 2 Measurement and Signal
Converters (Chapters 11-12)

AMCP 706-139 Section 4 Power Elements and System
Design (Chapters 14-20)

An index for the material in all four sections is placed at the end of Section 4.

Elements of the U. S. Army Materiel Command having need for handbooks may submit requisitions or official requests directly to Publications and Reproduction Agency, Letterkenny Army Depot, Chambersburg, Pennsylvania 17201. Contractors should submit such requisitions or requests to their contracting officers.

Comments and suggestions on this handbook are welcome and should be addressed to Army Research Office-Durham, Box CM, Duke Station, Durham, North Carolina 27706.

CHAPTER 13

AMPLIFIERS USED IN CONTROLLERS

13-1 ELECTRONIC AMPLIFIERS*

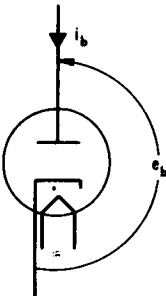
13-1.1 VACUUM TUBES

Vacuum tubes are thermionic devices in which a number of electrodes are contained inside a sealed envelope of glass, metal, or ceramic, which is evacuated to the lowest practical gas pressure. Types of tubes generally used in amplifiers include diodes, triodes, tetrodes, pentodes, and beam power tubes. Table 13-1 lists the symbols normally used with these types of tubes and their associated circuits.

13-1.2 Diodes

The simplest form of vacuum tube is a diode, which contains two electrodes (a plate

and a cathode) and a heater. In the symbolic representation of Fig. 13-1, the top element is the plate, the middle element is the cathode, and the heater is at the bottom. The



*By A. K. Suskind

Fig. 13-1 Symbolic representation of a diode.

TABLE 13-1 SYMBOLS

| | |
|---|----------|
| Control-grid supply voltage | E_{cc} |
| Plate supply voltage | E_{bb} |
| Instantaneous total grid voltage | e_c |
| Instantaneous total plate voltage | e_b |
| Instantaneous total plate current | i_b |
| Quiescent (zero excitation) value of plate voltage | E_{bo} |
| Quiescent (zero excitation) value of plate current | I_{bo} |
| Instantaneous value of alternating component of grid voltage | e_g |
| Instantaneous value of alternating component of plate voltage | e_p |
| Instantaneous value of alternating component of plate current | i_p |

heater, usually enclosed within the cathode, is used to raise the temperature of the cathode to the point where the cathode acts as a source of free electrons. The electrons leave the cathode and a great many of them travel to the plate, forming a unidirectional path for controllable current flow from the plate to the cathode (called *plate current*).

13-1.3 Control of electron flow. Useful control of electron flow from the cathode cannot be achieved by varying the heater potential. Therefore, heaters are always connected to sources of fixed potential, and heater symbols are frequently omitted from electronic-circuit diagrams.

In a diode, the only useful way to control electron flow is by control of the plate voltage e_b (the potential difference between the plate and cathode). The relationship between plate voltage e_b and plate current i_b for constant cathode temperature is shown in Fig. 13-2, where the curve is drawn for positive values of e_b . For all negative values of e_b , plate current i_b is zero. Reference directions (indicated by arrows in Fig. 13-1, for example) are chosen so that e_b is a positive number whenever the plate is actually positive with respect to the cathode, and i_b is a positive number when current flows from plate to cathode.

13-1.4 Diodes as rectifiers. The diode is a nonlinear device. One useful application of this nonlinearity is in a rectifying circuit such as Fig. 13-3. If the simplifying assumption is made that the e_b - i_b curve has a constant slope R_b in the conducting region, and infinite impedance in the nonconducting region, output current is given by

$$i_b = \frac{E_m}{R_b + R} \left(\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{3\pi} \cos 2\omega t - \frac{2}{15\pi} \cos 4\omega t - \dots \right) \quad (13-1)$$

The time-invariant (or d-c) component of this waveform can be isolated by adding a low-pass filter to the output. See Gray⁽¹⁾ for a more detailed discussion of rectifier circuits.

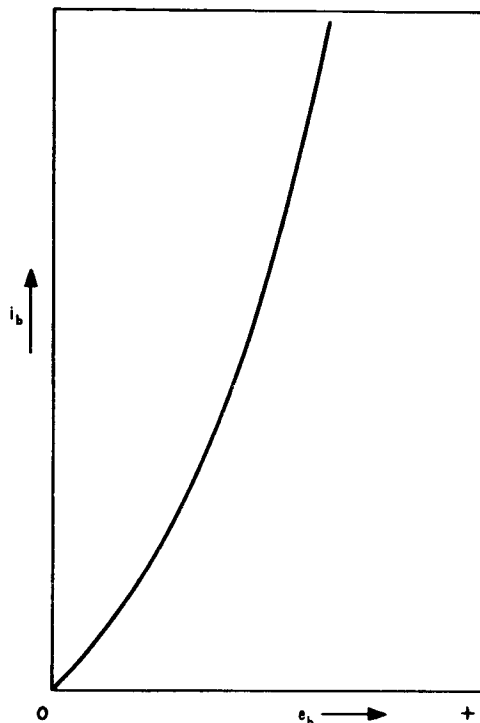


Fig. 13-2 Volt-ampere curve of a diode cathode temperature constant.

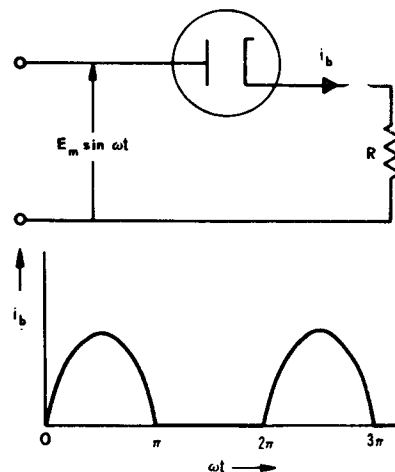


Fig. 13-3 Diode used as rectifier.

13-1.5 Triodes

If a third electrode, called the *control grid*, is inserted between the plate and cathode of a diode, the tube becomes a triode. The control grid provides an additional means of controlling electron flow. Plate current is now a function of both the control-grid and plate voltages and, since a change in grid voltage can result in a larger change in plate voltage, amplification can occur. The triode is shown symbolically in Fig. 13-4.

13-1.6 Plate characteristics. The volt-ampere (e_b - i_b) characteristics of a triode comprise a family of curves with grid voltage e_c (grid-to-cathode voltage) a parameter. These curves are called *plate characteristics*; a typical family is shown in Fig. 13-5. These curves demonstrate that:

(a) Plate current can flow only when the plate is positive with respect to the cathode.

(b) For a given value of plate voltage e_b , increasingly negative values of grid voltage e_c decrease plate current i_b and, if e_c is made sufficiently negative, conduction ceases. This is called *cutoff*, and the value of e_c that causes cutoff is called the *cutoff voltage*.

(c) The curves are not straight lines.

(d) Curves for successive values of e_c are displaced by unequal amounts.

It follows that, even when the plate voltage is positive and the grid voltage is above cutoff, the triode behaves as a nonlinear device. For precise results, therefore, nonlinear analysis must be used.

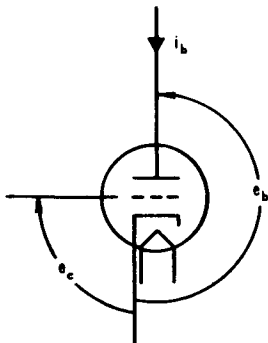


Fig. 13-4 Symbolic representation of a triode.

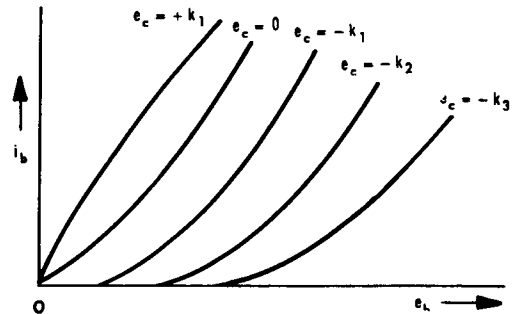


Fig. 13-5 Plate characteristics of a triode.

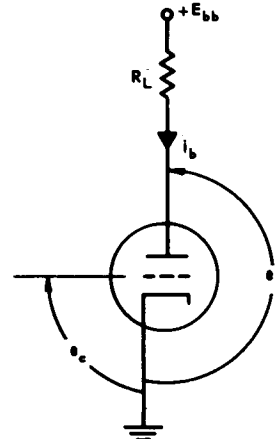


Fig. 13-6 Triode amplifier.

13-1.7 Graphical analysis. The most effective technique is graphical analysis. Using the typical triode amplifier in Fig. 13-6, which includes a load resistance R_L in its plate circuit, graphical analysis can be accomplished as follows:

(a) On the plate-voltage axis of the plate characteristics, mark the point corresponding to the plate supply voltage E_{bb} .

(b) From this point, draw a straight line with a slope $-1/R_L$ to the i_b axis. This sloping line is called the *load line*.

(c) For each value of e_c , determine the corresponding values of e_b and i_b using the

abscissa and ordinate, respectively, of the intersection of the load line with the e_c line of interest.

An example of this construction is given in Fig. 13-7. For a more complete discussion, see Preisman's book.⁽²⁾

13-1.8 Linear approximations. As an analysis tool, graphical construction is reasonably straightforward, particularly when the circuit associated with the tube is relatively simple. In many cases, however, complex configurations are involved and graphical analysis becomes tedious. It is especially so when nonresistive elements are encountered. It is often possible to make a good linear approximation of the tube characteristics, if the region of operation on the plate characteristics is small enough, so that

(a) The slope of the plate characteristics is approximately constant over that region.

(b) The plate characteristics are almost uniformly spaced over that region.

13-1.9 Region of operation. The region of operation is determined by the choice of E_{bb} , R_L , and the range over which e_c varies due to the input signal. In amplifiers for servomechanisms, it is usually possible (in all but the last stage) to select E_{bb} , R_L , and the tube type so that the grid is never operated positively or in the nonlinear region near cutoff. Under these conditions, the assumption of linear behavior usually leads to satisfactory results.

13-1.10 Linear equivalent circuits. The dynamic voltage-source linear equivalent circuit for a triode is given in Fig. 13-8. Here, e_g , e_p , and i_p are time-varying quantities and the grid-to-cathode impedance is shown as infinite. This is primarily justified by the assumption that the grid is never allowed to become positive with respect to the cathode and, as a result, no significant grid current flows. In the equivalent circuit, the amplification factor μ is defined as

$$\mu = - \left. \frac{\partial e_b}{\partial e_c} \right|_{i_b = \text{constant}} \quad (13-2)$$

The factor μ can be computed from the plate characteristics of Fig. 13-5 by finding (at the

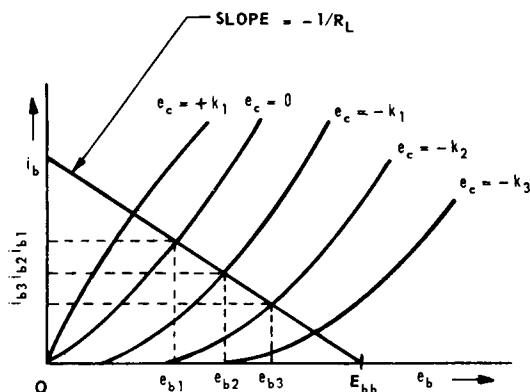


Fig. 13-7 Graphical analysis of a triode amplifier.

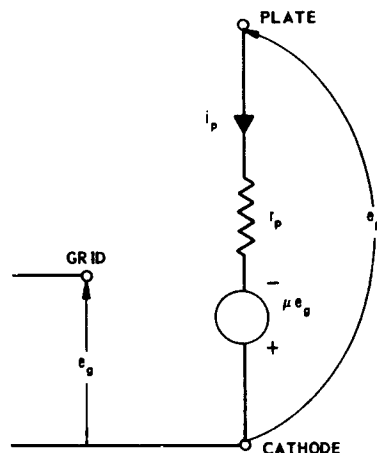


Fig. 13-8 Linear equivalent circuit of a triode.

approximate center of the operating region, and along a horizontal line) the change in plate voltage due to a change in grid voltage and then forming the quotient of these two variables. The plate resistance r_p , in ohms, is defined as

$$r_p = \left. \frac{\partial e_b}{\partial i_b} \right|_{e_c = \text{constant}} \quad (13-3)$$

The factor r_p can be computed by measuring the slope of an $e_b - i_b$ curve at the approximate center of the operating region. A third parameter, the transconductance, is also convenient to use at times. Transconductance, in mhos, is defined as the change in plate current per unit change in control-grid voltage and is given by

$$g_m = \left. \frac{\partial i_b}{\partial e_c} \right|_{e_b = \text{constant}} = \frac{\mu}{r_p} \quad (13-4)$$

For different types of triodes, typical values of μ range from 2 to 100, r_p from 1000 to 20,000 ohms, and g_m from 1000 to 5000 micromhos (1×10^{-6} mho = 1 micromho).

13-1.11 Alternate linear equivalent circuit.

An alternate linear equivalent circuit is shown in Fig. 13-9. In this circuit, a current source $g_m e_g$ and a parallel resistance r_p are used instead of the voltage source μe_g in series with r_p as shown in Fig. 13-8. The choice of which equivalent circuit to use in an analysis is entirely a matter of convenience because the two circuits shown give identical results.

13-1.12 Quiescent operating point. As indicated above, the linear model usually applies only to variations in e_b and i_b from the quiescent operating point, which is the point $e_b = E_{b0}$, $i_b = I_{b0}$, corresponding to zero input signal. The input signal e_g is usually superimposed on a negative bias E_{cc} , so that positive

values of e_g do not result in positive values of e_c . It follows that the quiescent operating point is found at the point where the zero-signal load line intersects the $e_c = E_{cc}$ characteristic. This point should be determined from the plate characteristics.

13-1.13 Phase shift. The zero-signal load is usually simple to determine because it takes into consideration only the resistive components of the load (capacitors being considered as open circuits and inductances as short circuits). Variations in e_b and i_b due to input signals are denoted as e_p and i_p . Using these definitions, the following may be written:

$$e_b = E_{b0} + e_p \quad (13-5)$$

$$i_b = I_{b0} + i_p \quad (13-6)$$

$$e_c = E_{cc} + e_g \quad (13-7)$$

The signs of e_p and i_p may be negative, but the signs of e_b and i_b are not. In fact, Figs. 13-8 and 13-9 show that e_p is negative when e_g is positive, and e_p is positive when e_g is negative. Thus, a triode causes a phase shift of 180° between grid and plate.

13-1.14 Pentodes and Beam-Power Tubes

Among vacuum tubes containing more than three electrodes, beam-power tubes and pentodes are of primary importance in servomechanism applications. A beam-power tube contains a cathode, a control grid, a screen grid, beam-forming electrodes, and a plate. A pentode contains a cathode, a control grid, a screen grid, a suppressor grid, and a plate. These two tube types are shown symbolically in Fig. 13-10. For servomechanism applications, the significant change brought about by the addition of the screen grid is the ability to achieve higher gain in beam-power and pentode circuits than is possible in triode circuits. However, since the screen grid is usually operated at a fixed d-c voltage (relative to the cathode), which is intermediate between plate supply and cathode potentials, an additional power supply is often required. This frequently outweighs the advantage of higher gain when over-all amplifier complexity is considered. In this respect, the triode is

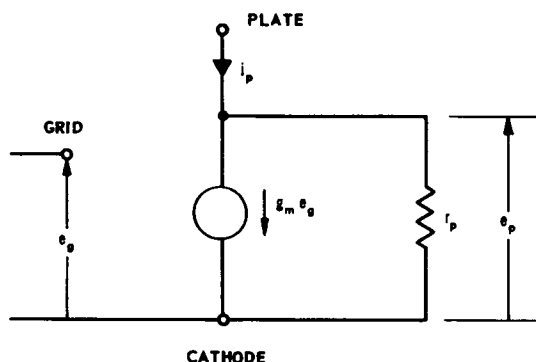


Fig. 13-9 Alternate equivalent circuit of a triode.

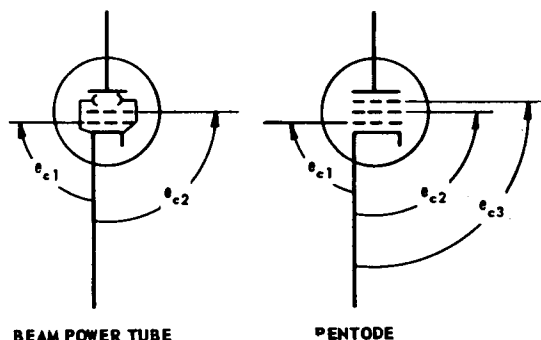


Fig. 13-10 Symbolic representation of a beam-power tube and a pentode.

superior to multigrid tubes because it requires no screen supply. It is often simpler to add more stages of amplification than to provide additional power supplies. Another feature that triodes possess is greater linearity of operation with large signal-voltage inputs.

13-1.15 Plate characteristics. Figure 13-11 shows typical plate characteristics of a beam-power tube, while those for a pentode are shown in Fig. 13-12. Both figures show that the plate current changes very little with changes in plate voltage when operation is above the knee of the curves and when all grid voltages are constant. It follows, therefore, that the plate resistance r_p of these tubes is very high, usually ranging from about 100,000 ohms to 10 megohms, as compared with a triode r_p of about 1000 to 20,000 ohms. Since triode transconductance is approximately equal to pentode transconductance, when similar construction is used, it follows that the amplification factor μ , given by

$$\mu = g_m r_p \quad (13-8)$$

is much higher for multigrid tubes than for triodes.

13-1.16 Linear equivalent circuits. When screen-grid and suppressor-grid voltages are constant, and operation is above the knee of the curves, the linear equivalent circuits of a multigrid tube have the same form as the

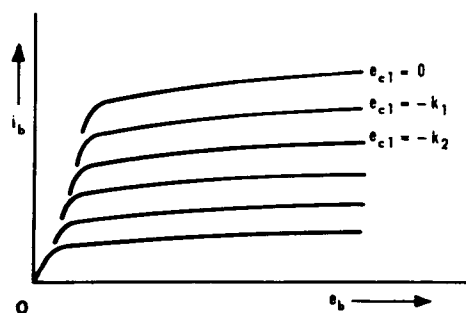


Fig. 13-11 Plate characteristics of a beam-power tube with constant screen voltage.

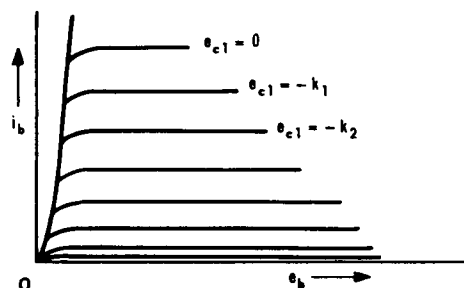


Fig. 13-12 Plate characteristics of a pentode with constant suppressor and screen voltages.

triode linear equivalent circuits of Figs. 13-8 and 13-9. For analysis of pentode operation, Fig. 13-9 is usually more convenient. When r_p is high compared with any load impedance connected to the plate, it is possible to consider the plate resistance as infinite. In Fig. 13-9, the tube can then be considered simply as a current source $g_m e_g$.

13-1.17 Graphical analysis. Graphical analysis of multigrid tubes is similar to that of triodes.

13-1.18 Interelectrode Capacitance

In the linear equivalent circuits heretofore discussed, no interelectrode capacitances were shown. These capacitances actually exist, but their values are so small (a few micromicrofarads) that they are of no consequence at

the signal frequencies encountered in servomechanisms. Therefore, interelectrode capacitance may be ignored, except when amplifier stability is considered.

13-1.19 Tube Specifications

In addition to the plate characteristics and the parameter values for the equivalent circuit, there are other important tube specifications, listed in tube-manufacturers' catalogs, which the circuit designer must take into consideration. While these additional specifications do not affect the theoretical operation, they do represent application limits. If these limits are exceeded, tube performance will differ from the given characteristics, or the period of operation over which the tube parameters remain at approximately the published values will be materially narrowed. These additional specifications must therefore be ascertained and adhered to. Of particular interest is the average plate dissipation P_p , which is defined as the time average (over a cycle) of the product of total plate voltage and total plate current. For sinusoidal operation

$$P_p = \frac{1}{2\pi} \int_0^{2\pi} e_p i_p d(\omega t) \quad (13-9)$$

For linear operation, the plate dissipation is greatest under quiescent conditions and is given by

$$P'_p = E_{b0} I_{b0} \quad (13-10)$$

where E_{b0} and I_{b0} are the plate voltage and plate current, with zero input signal. Since servo amplifiers usually receive no error signals for extended periods of time, quiescent-condition plate dissipation frequently determines whether the dissipation rating is being exceeded. The power dissipation of the screen is determined in a similar manner; i.e., by computing the time average of the product of screen voltage and total screen current.

13-1.20 LINEAR ANALYSIS OF SIGNAL-STAGE VACUUM-TUBE VOLTAGE AMPLIFIERS

In servomechanisms, voltage amplifiers are used primarily to increase the level of the control signal.

13-1.21 Characteristics of Tubes Used

The tubes used in low-level voltage amplifiers have the following characteristics:

(a) Triodes have high values of μ (e.g., 100) and pentodes have high values of g_m (e.g., 4000 micromhos).

(b) Maximum current ratings are moderate (e.g., 10 ma).

(c) Plate-dissipation ratings are moderate (e.g., 1 watt).

13-1.22 Simple Amplifier

The simplest voltage-amplifier circuit is that of Fig. 13-6, the linear equivalent of which is shown in Fig. 13-13. The gain of such an amplifier is defined as the ratio of the variational component of the output voltage to the variational component of the input voltage. The mathematical expression for this gain is

$$K = \frac{e_o}{e_{in}} = \frac{-\mu R_L}{R_L + r_p} \quad (13-11)$$

(The minus sign results from the 180° phase relationship between e_o and e_{in} .) In designing amplifiers, the usual practice is to attempt to achieve the required gain with the least number of stages of amplification. This naturally leads to an attempt to realize the maximum

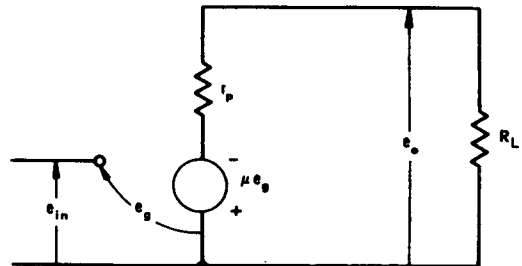


Fig. 13-13 Equivalent circuit of a simple plate-loaded amplifier

gain from each stage. Equation (13-11) indicates that the maximum possible gain equals μ and can be attained by making R_L much larger than r_p . To prevent this large value of R_L from producing plate-current values so low as to make r_p large (r_p is inversely related to i_b), a high plate supply voltage is required, reaching a value that is usually impractical. Practical values of E_{bb} do not exceed a few hundred volts and usually restrict the gain to a level that is only 60 to 80 percent of the amplification factor. Similar restrictions apply to a pentode, since g_m is approximately proportional to i_b .

13-1.23 Series Tube Triode Amplifier

In one method of attempting to realize the maximum possible gain, R_L is replaced by a second vacuum-tube circuit, as in the series tube triode amplifier of Fig. 13-14. Tube V2 and its cathode resistor R_k are approximately the equivalent of a battery of $\mu_2 E_{cc}$ volts plus a series resistor of value $r_{p2} + R_k(1 + \mu_2)$. As an example, consider a series tube circuit using both sections of a Type 6SL7 dual-triode, where

$$\mu_1 = \mu_2 = 70$$

$$r_{p1} = r_{p2} = 44,000 \text{ ohms}$$

$$R_k = 22,000 \text{ ohms}$$

$$E_{bb} = 350 \text{ volts}$$

$$E_{cc} = 45 \text{ volts}$$

The equivalent plate supply voltage is

$$E'_{bb} = E_{bb} + \mu_2 E_{cc} = 350 + (70 \times 45) = 3500 \text{ volts}$$

the equivalent plate load on V2 is

$$R'_L = r_{p2} + (1 + \mu_2) R_k = 44,000 + (71 \times 22,000) \approx 1.6 \text{ megohms}$$

and the gain is

$$K = \frac{e_o}{e_{in}} \approx \mu = 70$$

The series tube amplifier has one drawback, however, that makes it impractical except in rare instances. This drawback is the necessity for a floating supply E_{cc} .

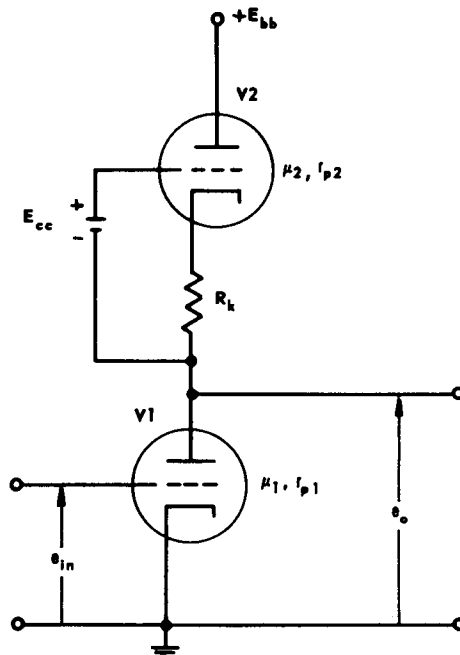


Fig. 13-14 Series tube amplifier.

13-1.24 Cascode Amplifier

A more practical circuit for realizing maximum possible gain is the cascode triode amplifier of Fig. 13-15. With tubes V1 and V2 identical, the gain of this circuit is

$$K = \frac{e_o}{e_{in}} = \frac{-\mu (\mu + 1) R_L}{(\mu + 2) r_p + R_L} \quad (13-12)$$

which reduces to

$$K = -g_m R_L \quad (13-13)$$

when

$$(\mu + 2) r_p \gg R_L \text{ and } \mu \gg 1$$

This demonstrates that the gain of a cascode triode amplifier equals that of a single pentode amplifier with infinite r_p . However, the pentode amplifier requires a screen-bias supply and the triode cascode amplifier does not. On the other hand, the disadvantage of the cascode triode amplifier is the necessity for a second tube. If two tubes are to be used, the second tube could be more advantageously

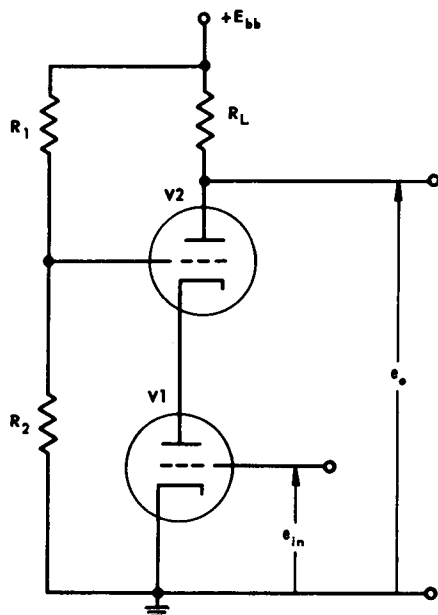


Fig. 13-15 Cascode amplifier circuit.

employed as a second voltage amplifier in cascade with the first, in which case an even larger over-all gain than either the cascode or pentode amplifier would produce might be realized. Cascode circuits are used primarily in direct-coupled amplifiers where the problem of cascading amplifier stages makes it desirable to maximize the gain per stage, even at the expense of extra tubes.

13-1.25 Cathode Followers

In equivalent circuits, the symbol e_g represents the variational component of grid-to-cathode voltage. When the tube circuit has internal feedback, e_g does not equal the externally applied input voltage. For example, in Fig. 13-16, which is the circuit schematic of a cathode follower, e_g is expressed as

$$e_g = e_{in} - e_k$$

where e_{in} is the time-varying component of the input voltage and e_k is the time-varying component of the cathode-to-ground voltage.

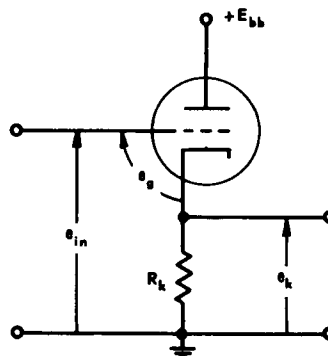


Fig. 13-16 Cathode follower.

Successive manipulations of the linear equivalent circuit (Fig. 13-17) demonstrate the nature of e_g . When the generator voltage μe_k (which has the same polarity as e_k) is considered as due to a fictitious resistor μR_k and is then lumped with R_k in the final step of Fig. 13-17, the identity of the cathode terminal vanishes. Therefore, additional cathode loading cannot be applied to the final equivalent circuit of Fig. 13-17. The equations for e_k and gain K are

$$e_k = i_p R_k = \frac{\mu e_{in} R_k}{R_k (1 + \mu) + r_p} \quad (13-14)$$

$$K = \frac{e_k}{e_{in}} = \frac{\mu R_k}{R_k (1 + \mu) + r_p} = \frac{\mu R_k}{(1 + \mu) \left(R_k + \frac{r_p}{1 + \mu} \right)} \quad (13-15)$$

The voltage gain of a cathode follower is always less than unity, no matter how large R_k is made. To a load connected from cathode to ground, the cathode follower appears as a generator with an open-circuit voltage as

AMPLIFICATION

given by Eq. (13-14), and the variational output impedance R_o is given by

$$R_o = \frac{R_k \left(\frac{r_p}{1 + \mu} \right)}{R_k + \left(\frac{r_p}{1 + \mu} \right)} \approx \frac{r_p}{1 + \mu} \approx \frac{1}{g_m} \quad (13-16)$$

if

$$R_k \gg \frac{r_p}{1 + \mu} \text{ and } \mu \gg 1$$

Practical values of R_o range from 100 to 1000 ohms.

The cathode follower is nearly always used to couple a low-impedance load to a high-impedance signal source, thereby reducing the severe loading effect (on the source) that is imposed by other types of coupling circuits. Where even lower output impedances are desired, the White cathode follower, comprising two tubes interconnected as shown in Fig. 13-18, can be used. The equivalent circuit (when V1 and V2 are identical) is shown in Fig. 13-19. This configuration can result in practical circuits with an output impedance of 10 to 100 ohms and with a gain approximately the same as that for a single-tube cathode follower.

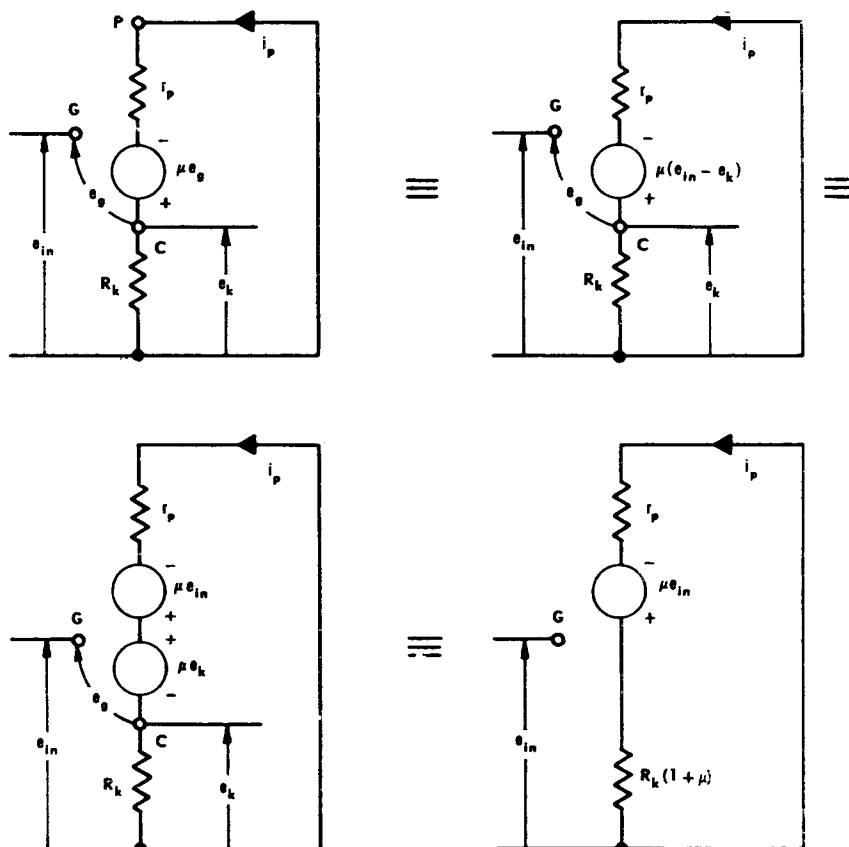


Fig. 13-17 Equivalent circuit of a cathode follower and its manipulation.

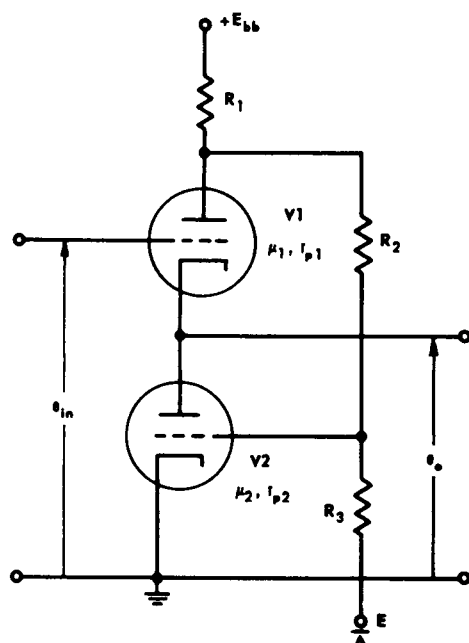


Fig. 13-18 White cathode follower.

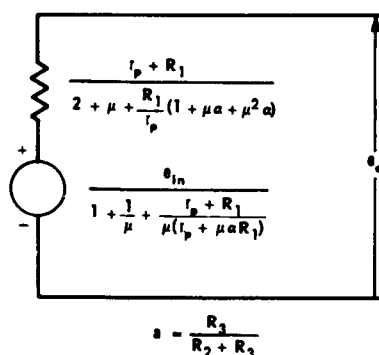


Fig. 13-19 Thevenin equivalent circuit of White cathode follower.

Where the voltage gain of a cathode follower must approach unity as closely as possible, the resistor R_k in Fig. 13-16 can be replaced by a second tube as in Fig. 13-20. The gain expression then becomes

$$K = \frac{e_o}{e_{in}} = \frac{\mu_1 [r_{p2} + R_k (1 + \mu_2)]}{(1 + \mu_1) \left[r_{p2} + R_k (1 + \mu_2) + \frac{r_{p1}}{1 + \mu_1} \right]} \approx 1 \quad (13-17)$$

when

$$r_{p2} + R_k (1 + \mu_2) \gg \frac{r_{p1}}{1 + \mu_1} \text{ and } \mu_1 \gg 1$$

These inequalities can exist in a practical circuit without reducing the V1 plate current to such a small value that r_{p1} is large.

13-1.26 Simple Feedback Amplifier

Another common form of voltage amplifier is the circuit of Fig. 13-21, which is evolved by adding a feedback resistor R_k to the circuit of Fig. 13-6. The grid-to-plate gain expression for this circuit is

$$K = \frac{-\mu R_L}{R_L + r_p + R_k (1 + \mu)} \quad (13-18)$$

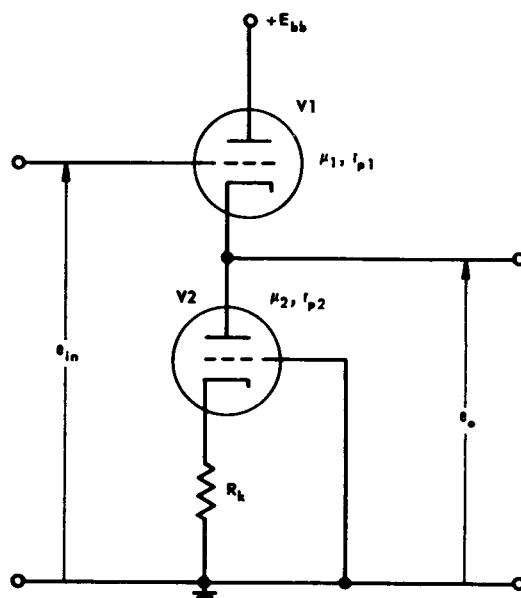


Fig. 13-20 Two-tube cathode follower.

and the variational output impedance that the load sees is

$$R_o = r_p + R_k(1 + \mu) \quad (13-19)$$

Because of the appearance of the term $R_k(1 + \mu)$ and also because of the approximate constancy of μ over a range of operating conditions (also during aging of the tube), the addition of R_k increases the gain stability in the presence of variations in r_p . It also lowers the gain and increases the effective output resistance. Output voltage e_o is expressed as

$$e_o = \frac{-\mu R e_{in}}{2R + \mu R + r_p} = -e_k \quad (13-20)$$

when

$$R_L = R_k = R$$

The circuit can be used as a phase splitter; i.e., a circuit with two outputs equal in magnitude but 180° out of phase. It can also be used as a coupling circuit between a single-ended input and a push-pull stage. The two output voltages must always be less than the input voltage.

13-1.27 Differential Amplifiers

Figure 13-22 shows the general form of a commonly used class of amplifiers called differential amplifiers. The time-varying components of the output voltages are given by the expressions

$$e_{o1} = -R_{L1} \left\{ \frac{-e_{in1}\mu_1[R_k(1 + \mu_2) + R_{L2} + r_{p2}] + e_{in2}\mu_2 R_k(1 + \mu_1)}{D} \right\} \quad (13-21)$$

$$e_{o2} = -R_{L2} \left\{ \frac{e_{in1}\mu_1 R_k(1 + \mu_2) - e_{in2}\mu_2[R_k(1 + \mu_1) + R_{L1} + r_{p1}]}{D} \right\} \quad (13-22)$$

where

$$D = (1 + \mu_1)(1 + \mu_2)R_k^2 - [(1 + \mu_1)R_k + R_{L1} + r_{p1}][(1 + \mu_2)R_k + R_{L2} + r_{p2}]$$

If both tubes are identical, $R_{L1} = R_{L2} = R_L$, and the output is taken between the plate of V1 and the plate of V2, then

$$e_o = e_{o1} - e_{o2} = \frac{\mu R_L}{R_L + r_p}(e_{in2} - e_{in1}) \quad (13-23)$$

This equation shows that the plate-to-plate output voltage of a differential amplifier is proportional to the *difference* between the input signals. If the output is taken between the plate of one tube and a point of fixed potential (such as ground), the output can still be made approximately proportional to the difference between the input signals by making

$$R_k(1 + \mu) \gg R_L + r_p$$

Then

$$e'_{o1} = \frac{-\mu R_L}{2(R_L + r_p)}(e_{in1} - e_{in2}) \quad (13-24)$$

$$e'_{o2} = \frac{\mu R_L}{2(R_L + r_p)}(e_{in1} - e_{in2}) \quad (13-25)$$

Because it produces an output signal that is either positive or negative, and proportional to the difference between the two input signals, this circuit is useful as a summing device in a servomechanism. If one of the grids is grounded, one output of the differential amplifier is in phase with the input, and the other output is 180° out of phase with the input. For this reason, the differential amplifier may be used as a phase splitter. It is then necessary that the two output voltages be of

equal magnitude. For this case it is sufficient, but not necessary, that

$$R_k(1 + \mu) \gg R_L + r_p$$

It is only necessary that

$$R_{L1} = \frac{R_k R_{L2}}{R_k + \frac{r_p + R_{L2}}{1 + \mu}}$$

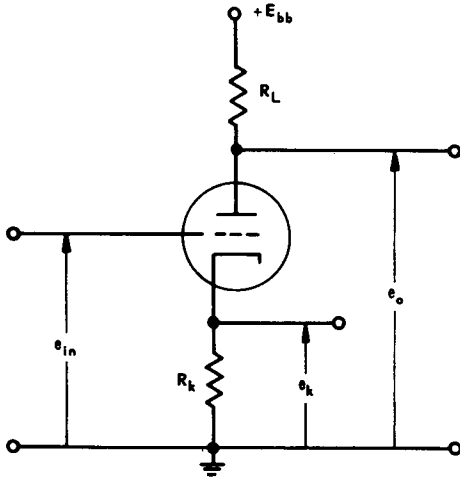


Fig. 13-21 Plate-and-cathode-loaded amplifier.

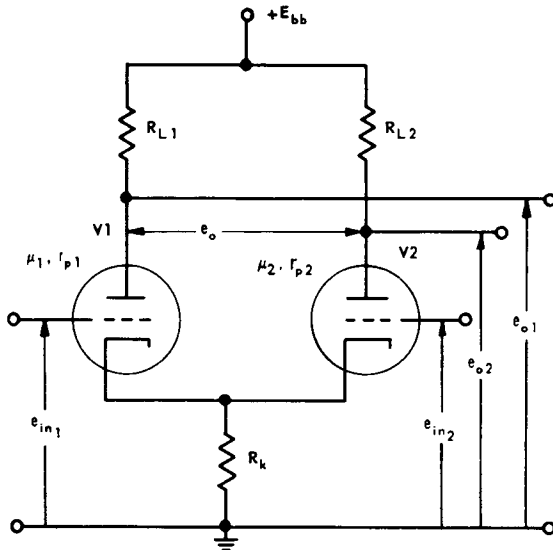


Fig. 13-22 Differential amplifier.

Because the output of a differential amplifier depends almost entirely upon the difference between the two input signals, the circuit is insensitive to equal changes in both signals. Thus, equal variations in the input signals, such as drift in the reference potentials to which they are returned, result in small output-voltage variations. This property makes differential amplifiers very useful for d-c amplifier applications.

13-1.28 Use of Pentodes

In all circuits previously discussed where a resistor R_k is connected in series with the cathode, the use of pentode tubes is not convenient. This is because maintenance of proper screen voltage (which must be constant with respect to the cathode) would require a separate power supply for the screen of each pentode stage, with the supply connected between screen and cathode. However, when signal frequencies are high enough (at least 10 cps), pentodes can be used without a separate or floating screen supply for each stage. This is accomplished by connecting all pentode screens to a single source of positive voltage through individual resistors R_s and by connecting a large capacitor (typically, 0.1 to 1 microfarad) between the screen and cathode of each pentode.

13-1.29 POWER AMPLIFIERS

In the previous discussion of voltage-amplifier circuits, no attention was paid to the current levels in the amplifier load because the power supplied to that load is of little consequence. However, in the final stage of a servo amplifier, the amplifier load is the output member of the servomechanism. Since the load is usually a motor or a similar power-consuming device, the final amplifier stage must be a power amplifier with adequate output.

13-1.30 Tubes Used in Power Amplifiers

The tubes used for power amplification differ from those used for voltage amplification. These differences include:

- (a) Greater maximum plate-dissipation ratings (e.g., 10 watts)

(b) Greater plate-current ratings (e.g., 200 ma)

(c) Greater range of grid voltage over which operation is linear (e.g., 15 volts)

Pentodes and beam-power tubes (e.g., Type 6L6) are commonly used as power-amplifier tubes because they produce a larger a-c load voltage than triodes do, for a given plate supply voltage and given maximum permissible harmonic distortion. In addition, pentodes are less likely than triodes to produce unwanted oscillations, when feeding inductive loads such as motors, because the grid-to-plate capacitance of a pentode is lower than that of a triode.

13-1.31 Push-Pull Power Amplifiers

In servomechanisms, the push-pull power amplifier is more generally used than the single-ended, or any other type of power amplifier. The push-pull circuit comprises two tubes of the same type connected as shown in Fig. 13-23. The two grid-signal voltages are of equal magnitude and 180° out of phase. The input transformer T1 can be replaced by a phase-splitting circuit as discussed in connection with Figs. 13-21 and 13-22. The major advantages of the push-pull circuit over the single-ended circuit (load connected to one tube or two tubes in parallel) can be summarized as follows:

(a) Higher plate-circuit efficiency (ratio of output power to power dissipated in tube). Efficiency can be as high as 65 percent in a push-pull amplifier, as against up to 30 percent for a single-ended amplifier. Hence, for a given pair of tubes dissipating a fixed amount of power, the push-pull circuit delivers a greater output.

(b) An output, or load, signal that is completely free of even harmonics. Tube nonlinearities introduce these harmonics, but push-pull circuit action cancels them. This permits operation of each tube with a bias that is high enough to cut off the tube during part of the negative grid-voltage half-cycle when there is an input sinusoid of maximum amplitude. Under these conditions, the higher values of plate-circuit efficiency (up to 65 percent) are achieved. Therefore, in servo amplifiers, E_{cc} is often adjusted for plate-current cutoff during some part of the negative portion of the maximum input sinusoid, but for less than 180° if self-bias is used.

(c) There is no net d-c component of plate current in the output-transformer primary. Therefore, the transformer may be lighter and smaller than one that is used to match a load to a single-ended power amplifier which must carry d-c current.

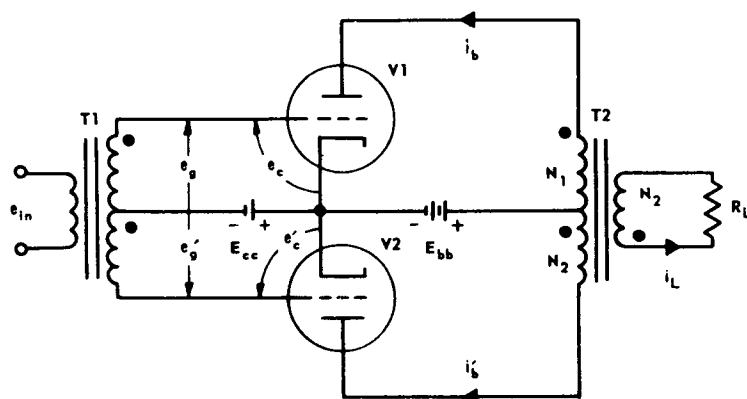


Fig. 13-23 Push-pull amplifier.

(d) Ripple in the plate supply voltage does not appear in the load, thus reducing the filtering requirements on the power supply.

13-1.32 Analysis of push-pull power amplifiers. Practical push-pull amplifiers must be analyzed by graphical techniques because they nearly always involve nonlinear tube operation. The analysis procedure is as follows:

(a) First, form the *composite* plate characteristics of the two tubes. This is done by arranging two sets of plate characteristics, as shown in Fig. 13-24A, so that the origin of the bottom set coincides with the point $e_b = 2E_{bb}$ of the top set.

(b) Add algebraically the two curves e_c , E_{cc} . In the example of Fig. 13-24B, E_{cc} is assumed to be equal to $-k_3$, and the dashed curve is the result of the addition. Relabel the resultant curve $e_g = 0$.

(c) Add algebraically the upper curve for $e_c = -k_2$ and the lower curve for $e'_c = -k_4$. Relabel the resultant curve $e_g = -k_2 + k_3 = k_1$.

(d) Add algebraically the upper curve for $e_c = -k_4$ and the lower curve for $e'_c = -k_2$. Relabel the resultant curve $e_g = -k_1$.

(e) Continue this procedure by combining upper e_c and lower e'_c until the complete set of composite characteristics is obtained (Fig. 13-24C).

(f) Through the point $e_b = E_{bb}$, draw a straight line with slope $-1/(N_1/N_2)^2 R_L$. Relabel the current axis $(N_2/N_1) i_L$ and the voltage axis $(N_1/N_2) e_L$. This is shown in Fig. 13-24D. Load voltage e_L and load current i_L may then be read off as a function of e_g .

13-1.33 Push-pull amplifiers with a-c supply. When the amplifier signal is an amplitude-modulated carrier, a power amplifier similar to the one shown in Fig. 13-25 is sometimes used. Its advantage over the conventional push-pull configuration is that it does not require a d-c power supply. The resulting signal waveforms are indicated in Fig. 13-26. The pulse-like nature of the plate-current waveform makes analysis exceedingly difficult,

particularly when pentodes are used. Therefore, circuit values for this configuration are more easily determined by experimental procedures.

13-1.34 Efficiency

In practice, vacuum-tube power amplifiers are used only where the load requirements are moderate (up to approximately 100 watts) because such amplifiers are relatively inefficient, with a maximum of about 65 percent.

13-1.35 CASCADING AMPLIFIER STAGES

When a single amplifier stage cannot provide the required gain, additional cascaded stages must be supplied until the requirement is met. In a cascaded amplifier, each successive stage amplifies the output of the preceding stage. There are two methods of cascading, each named according to the type of coupling circuit used between stages:

(a) Direct coupling, in which the total instantaneous value of the output of one stage affects the next stage.

(b) A-c coupling, in which only the time-varying component of the output of one stage affects the next stage.

13-1.36 Direct-Coupled Amplifiers

An example of a two-stage direct-coupled amplifier is given in Fig. 13-27. Assuming that $R_1 + R_2 \gg R_{L1}$, the gain of this amplifier is expressed as

$$K = \frac{\mu_1 \mu_2 R_{L1} R_{L2}}{(R_{L1} + r_{p1})(R_{L2} + r_{p2})} \times \frac{R_2}{R_1 + R_2} \quad (13-26)$$

The voltage divider R_1, R_2 reduces the gain but is needed to make the V2 grid negative with respect to its cathode. The voltage divider can be eliminated by coupling the plate of V1 to the grid of V2 by means of a battery E (Fig. 13-28). However, the separate battery makes this coupling method impractical from many viewpoints. Another way of eliminating the voltage divider is to couple the plate of V1 directly to the grid of V2. The cathode of V2 must then be returned to a positive-potential point, and R_{L2} to a still more positive point.

AMPLIFICATION

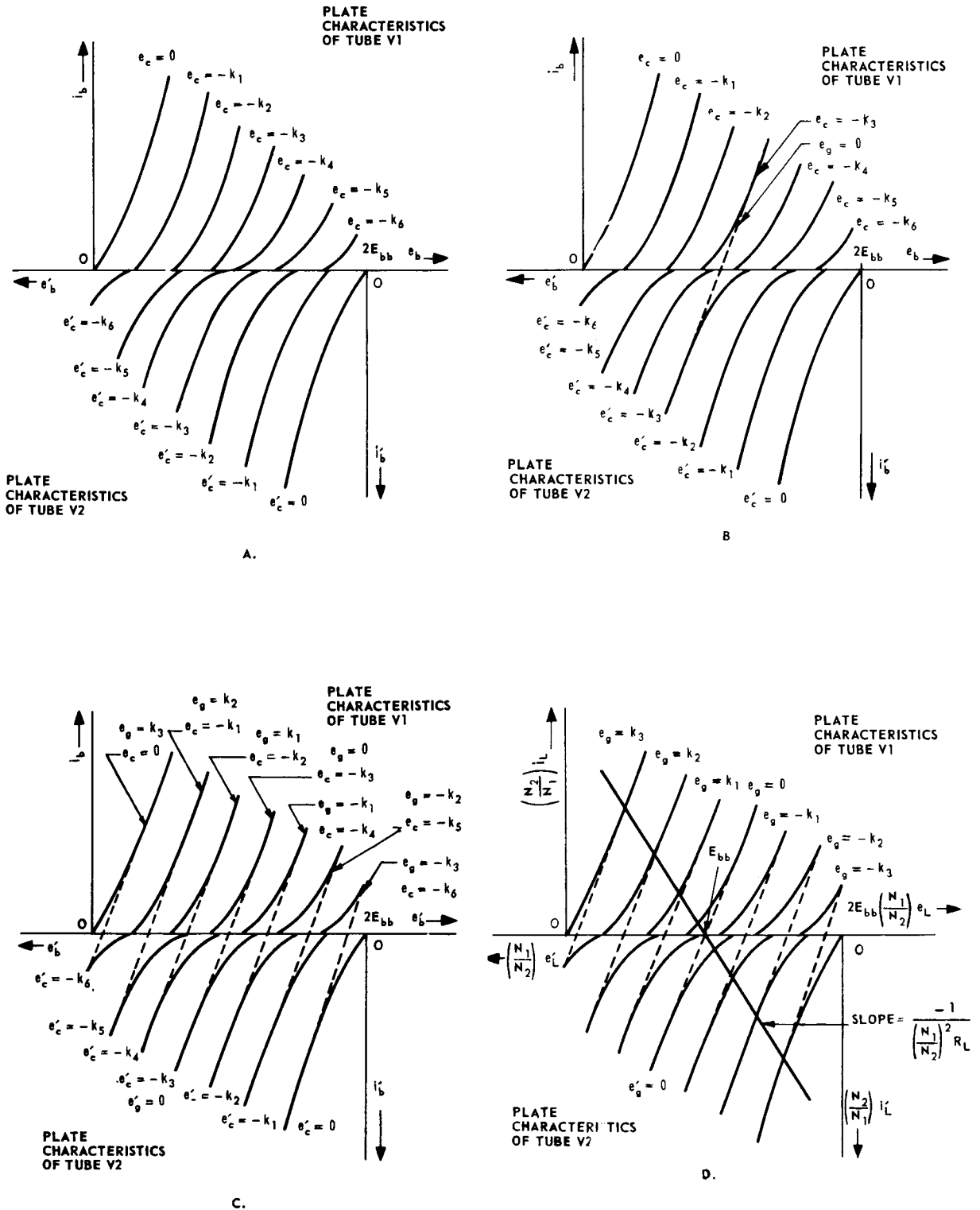


Fig. 13-24 Graphical analysis of a push-pull amplifier.

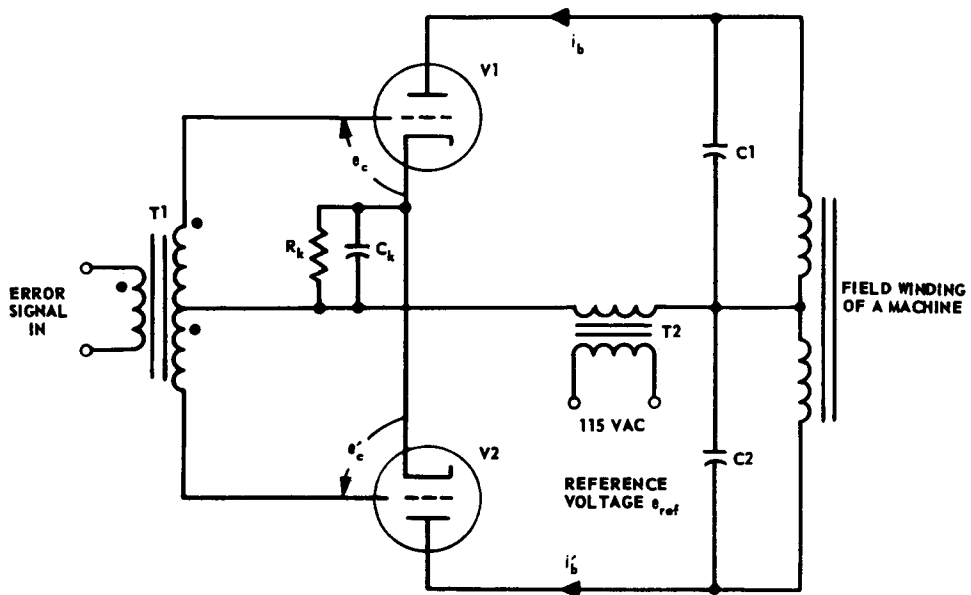


Fig. 13-25 Power amplifier with a-c supply.

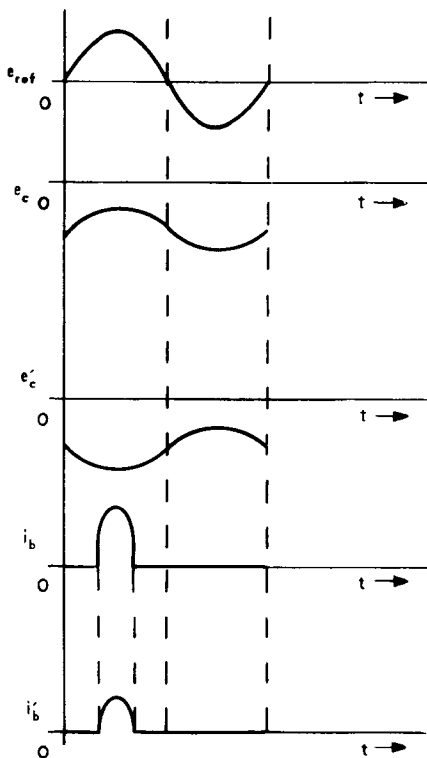


Fig. 13-26 Waveforms of power amplifier in Fig. 13-25.

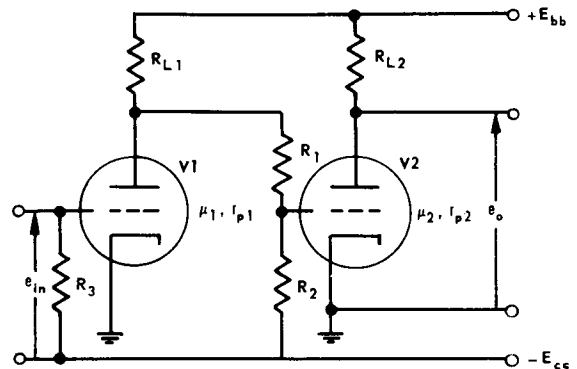


Fig. 13-27 Voltage-divider-coupled d-c amplifier.

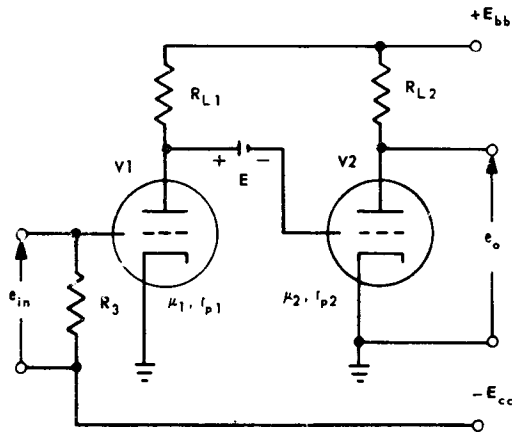


Fig. 13-28 Battery-coupled, d-c amplifier.

Four separate power supplies are therefore required (bias of V1, plate of V1, cathode of V2, and plate of V2). Again, an impractical situation arises. However, the need for four separate supplies can be obviated by using a single high-voltage supply and tapping off the required d-c voltages at appropriate points on a bleeder as shown in Fig. 13-29. Typical voltage levels corresponding to zero input signal are indicated on this figure. If the bleeder current is large compared with the tube currents, then the gain of the amplifier is

$$K = \frac{\mu_1 \mu_2 R_{L1} R_{L2}}{(R_{L1} + r_{p1})(R_{L2} + r_{p2})} \quad (13-27)$$

The heavy drain on the power supply, due to the large bleeder current, makes this method undesirable in many applications. Figure 13-30 shows still another method of eliminating the effect of the voltage divider. Resistance R_1 in Fig. 13-27 is replaced by a gas-discharge tube, such as a Type VR105 voltage regulator. A gas-discharge tube is characterized by a very nearly constant terminal voltage over a limited current range (e.g., 10 to 30 ma) so that it acts as an equivalent battery. Figure 13-30 is the simplest coupling circuit where signal levels are high enough so that the gas-discharge-tube noise (somewhat reduced by R_N and C_N) does not reduce the amplifier signal-to-noise ratio below acceptable limits. Iannone⁽³⁾ discusses in detail the use of gas-discharge tubes as coupling devices in d-c amplifiers.

13-1.37 Problems encountered in direct-coupled amplifiers. Not only are there practical difficulties in the design of coupling circuits in direct-coupled amplifiers, but the very nature of direct coupling makes the output level of such an amplifier dependent upon all amplifier components, tube characteristics, and power supply levels. Any variation

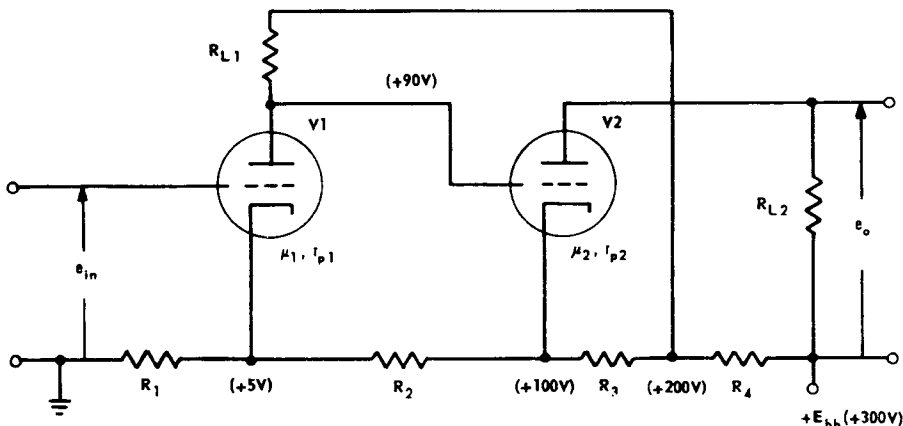


Fig. 13-29 D-c amplifier with single supply voltage.

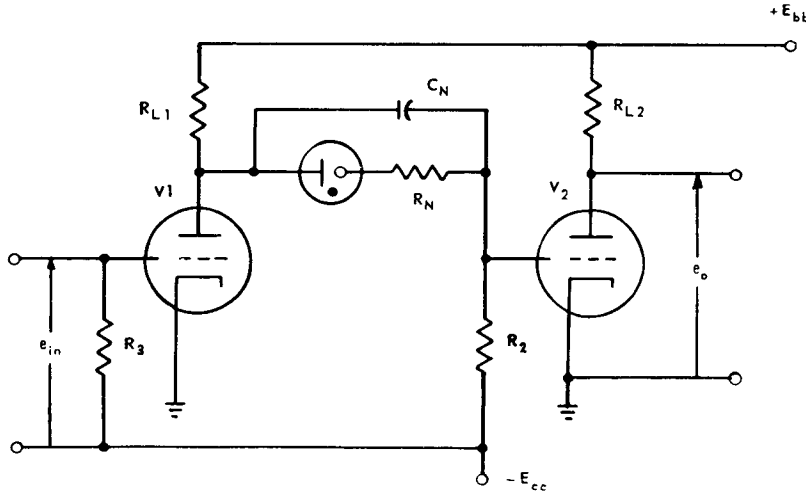


Fig. 13-30 D-c amplifier with gas-discharge tube coupling.

in these parameters (caused by aging of components, changes in environmental conditions, power supply ripple, etc.) changes the output level even though the input remains constant. The following measures are generally used to reduce variation (drift) in the output level:

(a) Use only components that are stabilized for age and changes in environmental conditions; e. g., tubes that have been "broken in" (operated for 100 to 200 hours before use) and temperature-stable metalized-type resistors.

(b) Incorporate good regulation of all power supply voltages, including filament voltage. (A change of 20 percent in filament voltage is approximately equivalent to a grid-voltage change of 0.2 volt. Over the nonlinear range of operation, changes in filament voltage also produce changes in r_p .)

In practice, these measures cannot be carried out to perfection. Tubes show the effects of aging, practical power supplies rarely have voltage regulation better than 0.1 percent, and good regulation of filament supplies often results in excessive weight and space requirements. The reduction of drift, therefore, requires the use of special circuit designs (par-

ticularly for the first stage of a direct-coupled amplifier) that inevitably lead to greater over-all amplifier complexity. One example of added complexity is the use of feedback, which can *reduce* the effect of drift but cannot *eliminate* it.

13-1.38 Drift-compensated direct-coupled amplifier. Figure 13-31 shows a circuit in which the output is independent of filament-voltage and supply-voltage variations, provided that the two tubes are identical. Assuming linear operation, the ratio of total output voltage to total input voltage is

$$\frac{e_o}{e_{in}} = \frac{-\mu R_3}{2r_p + R_3} \quad (13-28)$$

13-1.39 Bridge circuits. Bridge circuits, such as those in Figs. 13-32 and 13-33, eliminate drift due to changes in E_{bb} , filament voltage, and r_p . Assuming linear operation in Fig. 13-32, the ratio of output voltage to input voltage is

$$\frac{e_o}{e_{in}} = \frac{-\mu R_L}{r_p + 2R_k + 2R_L + R} \quad (13-29)$$

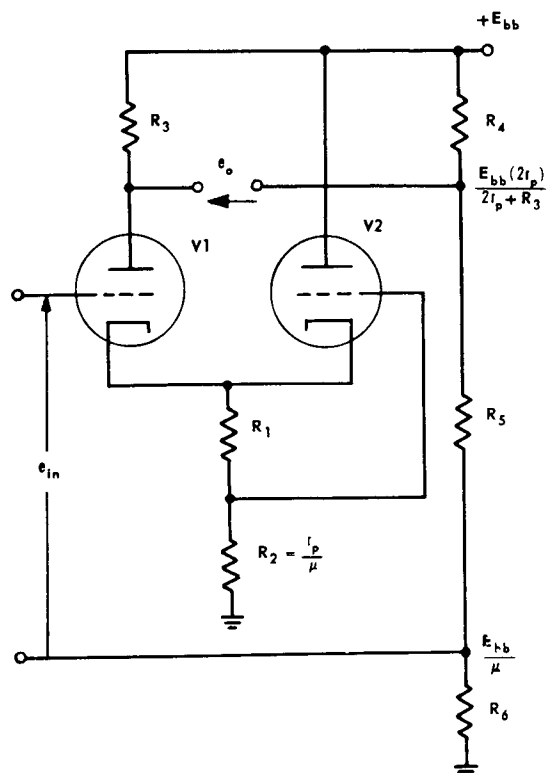


Fig. 13-31 Drift-compensated d-c amplifier.

and the output impedance R_o seen by the load is

$$R_o = R_k + \frac{r_p + R}{2} \quad (13-30)$$

Assuming linear operation in Fig. 13-33, a low output impedance circuit, it follows that

$$\frac{e_o}{e_{in}} = \frac{-\mu R_L}{r_p + R_k(1 + \mu) + R_L(2 + \mu) + R} \quad (13-31)$$

and

$$R_o = \frac{R_k(1 + \mu)}{2 + \mu} + \frac{r_p + R}{2 + \mu} \quad (13-32)$$

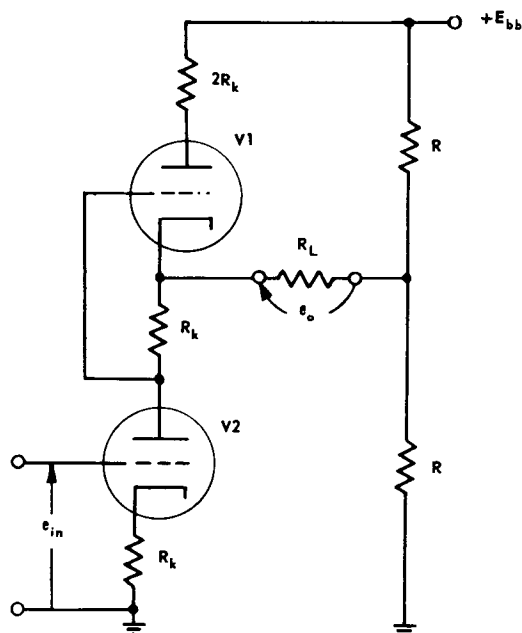


Fig. 13-32 Drift-compensated d-c amplifier.

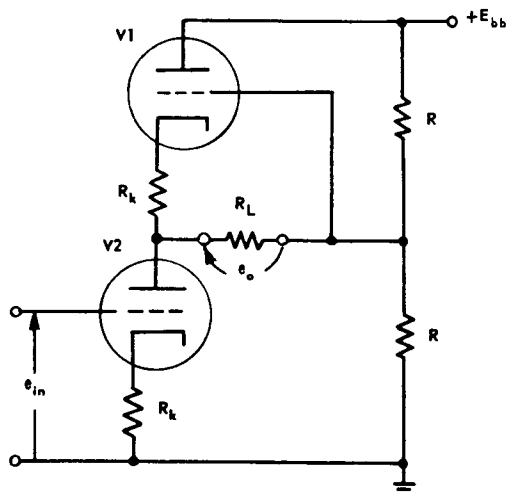


Fig. 13-33 Drift-compensated d-c amplifier.

13-1.40 A-C Coupled Amplifiers

Because the reduction of drift to a negligible amount often leads to considerable circuit complexity, direct-coupled amplifiers are generally avoided in servomechanism design. Instead, a-c coupled amplifiers are used. However, in an a-c coupled amplifier, low-frequency signal components are attenuated and, since typical signal frequencies in a servomechanism occur in the 0 to 20 cps band, the signal must first be modulated; i.e., superimposed on an a-c carrier. Typical carrier frequencies are 60 cps and 400 cps. The block diagram of a servomechanism incorporating an a-c coupled amplifier is shown in Fig. 13-34. If the amplifier load responds to modulated a-c signals, the demodulator is not used. An example of such a load is a 2-phase induction motor. If the input signal is already in the form of a modulated a-c signal, such as the output of a synchro control transformer, the modulator is not used. Servo compensation is shown ahead of the modulator in Fig. 13-34 because d-c compensating networks are simpler to design than a-c compensating networks.

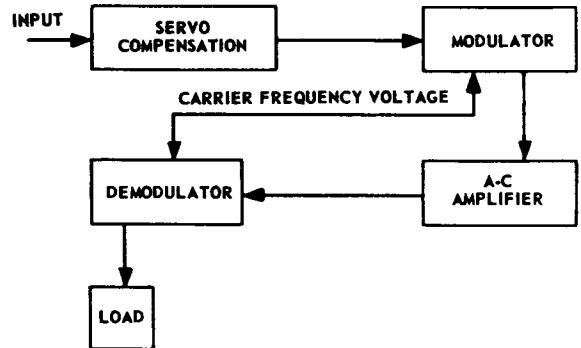


Fig. 13-34 Use of a-c amplifier to replace d-c amplifier.

13-1.41 Two-stage a-c coupled amplifier. Figure 13-35 is the circuit schematic of a simple two-stage a-c coupled amplifier. Resistors R_k and capacitors C_k furnish self-bias by raising the average cathode potential above ground. The grid circuit may then be returned to ground potential, as shown, and the need for a grid-bias supply is eliminated. C_k is large

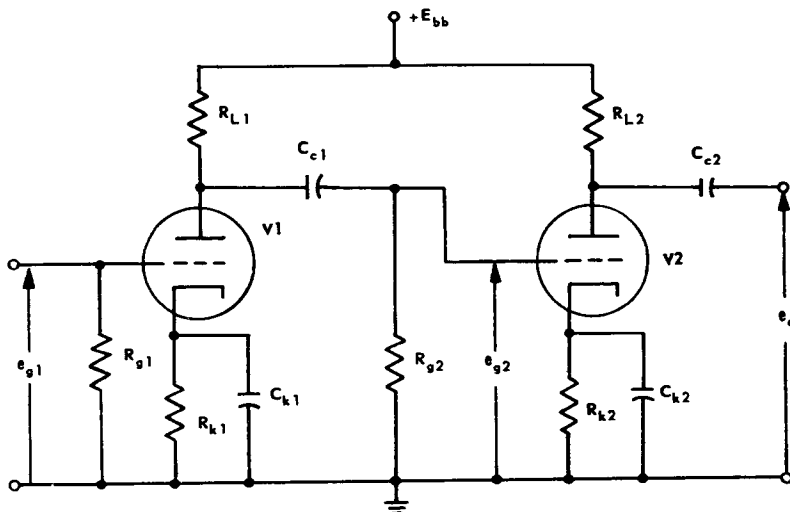


Fig. 13-35 Two-stage a-c amplifier (resistance-capacitance coupled).

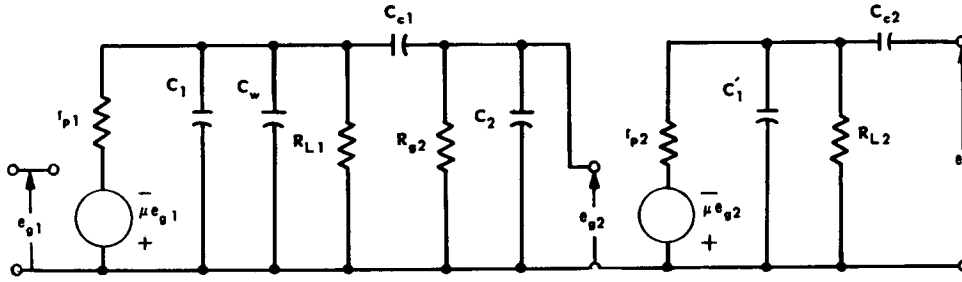
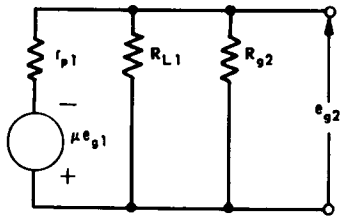
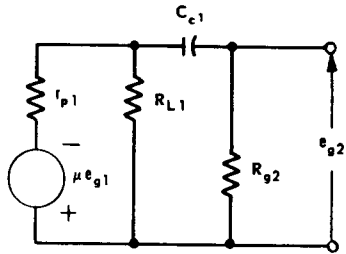


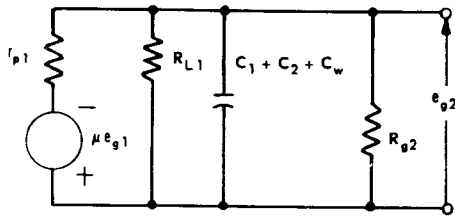
Fig. 13-36 Equivalent circuit of two-stage a-c amplifier in Fig. 13-35.



A. MID-FREQUENCY EQUIVALENT CIRCUIT



B. LOW-FREQUENCY EQUIVALENT CIRCUIT



C. HIGH-FREQUENCY EQUIVALENT CIRCUIT

Fig. 13-37 Simplified equivalent circuits of first stage of Fig. 13-35.

enough (1 to 40 microfarads) so that its reactance is negligible at the frequencies of interest, thereby eliminating the effect of R_k on dynamic response. For linear operation, the value of R_k is chosen so that

$$|E_{cc}| = I_{bo} R_k \quad (13-33)$$

where

$$|E_{cc}| = \text{absolute value of desired bias}$$

$$I_{bo} = \text{quiescent plate current}$$

The large capacity of C_k is provided in compact form by low-voltage electrolytic capacitors.

The equivalent circuit of the two-stage a-c coupled amplifier is shown in Fig. 13-36. For analysis, this equivalent circuit can be broken down into three separate simplified circuits as shown in Fig. 13-37. In these circuits, C_1 is the output capacitance of tube V1, C_2 is the input capacitance of tube V2, and C_w is the wiring capacitance. The gain expressions for the three frequency regions are

$$K_{mid} = \frac{e_{g2}}{e_{g1}} = \frac{-\mu R_{L1} R_{g2}}{r_{p1} R_{L1} + r_{p1} R_{g2} + R_{L1} R_{g2}} \quad (13-34)$$

$$K_{low} = \frac{K_{mid}}{1 - j \frac{f_1}{f}} \quad (13-35)$$

$$K_{high} = \frac{K_{mid}}{1 + j \frac{f}{f_2}} \quad (13-36)$$

where

f = frequency of input signal

f_1 = lower half-power frequency (response down to 3 db)

$$= \frac{r_{p1} + R_{L1}}{2\pi C_{c1}(r_{p1}R_{L1} + r_{p1}R_{g2} + R_{g2}R_{L1})}$$

f_2 = upper half-power frequency (response down 3 db)

$$= \frac{r_{p1}R_{L1} + r_{p1}R_{g2} + R_{L1}R_{g2}}{2\pi(C_1 + C_2 + C_w)r_{p1}R_{L1}R_{g2}}$$

Typical values of f_1 range from 1 to 20 cps, typical values of f_2 from 50 to 500 kc. Thus, in a servomechanism application, the high-frequency equivalent circuit needs consideration only when stability questions are involved. In a properly designed amplifier, it is usually permissible to assume that the gain is constant and given by the mid-frequency value K_{mid} . Plots of output voltage, gain, and phase shift as a function of frequency are shown in Fig. 13-38. When pentodes are used, and the effect of R_k and C_k is to be taken into account, then

$$\frac{\text{actual output voltage}}{\text{output voltage with fixed-bias operation}} = \frac{1}{1 + \frac{g_m R_k}{1 + j \frac{f}{f_3}}} \quad (13-37)$$

where

f = frequency of input signal

$$f_3 = \frac{1}{2\pi C_k R_k}$$

13-1.42 FEEDBACK AMPLIFIERS

13-1.43 Advantages

Negative feedback applied to an amplifier produces the following advantages:

(a) *Improved stability,*

$$\frac{dK'}{K'} = \frac{1}{1 - \beta K} \frac{dK}{K} \quad (13-38)$$

where

$\frac{dK'}{K'}$ = per unit change in gain with feedback

K = open-loop gain

β = fraction of output signal fed back

$\frac{dK}{K}$ = per unit change in gain without feedback

Since $|1 - \beta K| > 1$ for negative feedback, then

$$\frac{dK'}{K'} < \frac{dK}{K}$$

(b) *Modification of frequency response.* For example, addition of negative feedback to a single-stage a-c coupled amplifier (Fig. 13-39) decreases the lower half-power frequency and increases the upper half-power frequency. Thus,

$$f_{1f} = f_1 / (1 - \beta K_{mid}) \quad (13-39)$$

$$f_{2f} = f_2 (1 - \beta K_{mid}) \quad (13-40)$$

where

f_{1f} = lower half-power frequency with feedback

f_1 = lower half-power frequency without feedback

f_{2f} = upper half-power frequency with feedback

f_2 = upper half-power frequency without feedback

K_{mid} = mid-frequency gain of amplifier without feedback

(c) *Modification of input and output impedances.*

(1) For voltage feedback (signal proportional to output voltage fed back), the output impedance is given by

$$Z'_o = \frac{Z_o}{1 - \beta K_o} \quad (13-41)$$

where

Z'_o = output impedance with voltage feedback

Z_o = output impedance with input voltage to feed-forward section short-circuited

AMPLIFICATION

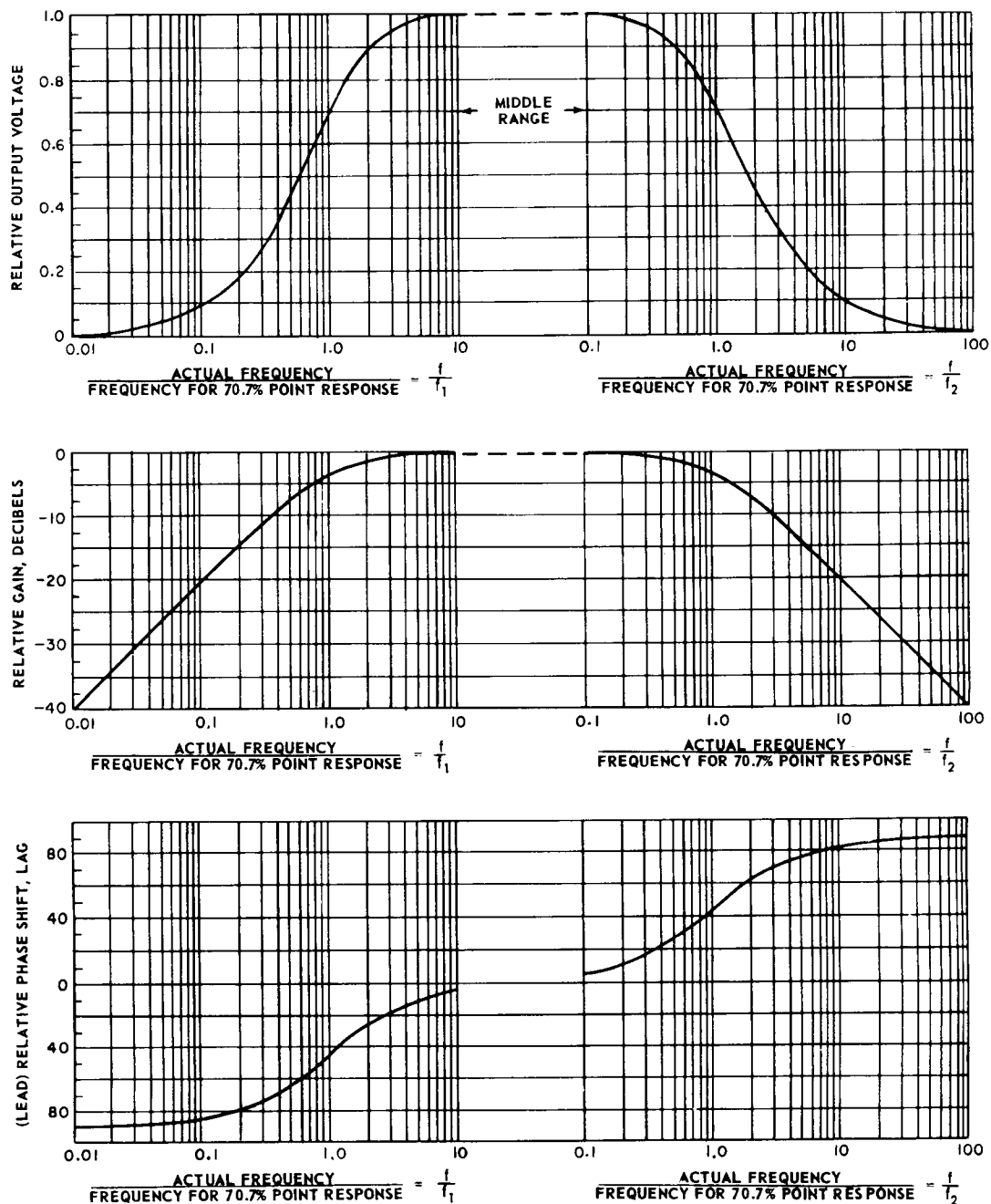


Fig. 13-38 Frequency response of single-stage a-c coupled amplifier using resistance-capacitance coupling.

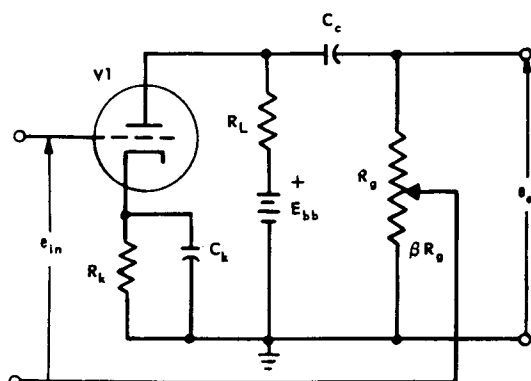


Fig. 13-39 Single-stage a-c coupled amplifier with voltage feedback.

β = fraction of output voltage fed back

K_o = open-loop gain of amplifier with load open-circuited

Since $|1 - \beta K_o| > 1$ for negative feedback, then

$$Z'_o < Z_o.$$

For example, consider the output impedance seen by R_k in the cathode follower of Fig. 13-16. Here,

$$Z_o = r_p$$

$$\beta = -1$$

$$K_o = \mu$$

$$Z'_o = \frac{r_p}{1 + \mu}$$

(2) For current feedback (signal proportional to output current fed back), the output impedance is given by

$$Z''_o = Z_o + Z_f(1 - K_o) \quad (13-42)$$

where

Z''_o = output-impedance with current feedback

Z_f = impedance across which feedback is developed

Since $|1 - K_o| > 1$ for negative feedback, then

$$Z''_o > Z_o$$

For example, consider the output impedance seen by the load resistor R_L in Fig. 13-21. Here,

$$Z_o = r_p + R_k$$

$$Z_f = R_k$$

$$K_o = -\mu$$

$$Z''_o = r_p + R_k(1 + \mu)$$

(3) When both voltage and current feedback are used, then

Z'''_o = effective output impedance with both types of feedback

$$= \frac{Z_o + Z_f(1 - K_o)}{1 - \beta K_o} \quad (13-43)$$

Techniques for finding impedance levels at any terminal pair are given in the book by Bode.⁽⁴⁾

(d) *Improvement in signal-to-noise ratio when noise is generated internally (not at input to amplifier).* The addition of feedback itself does not increase the signal-to-noise ratio. It does, however, permit an increase in the signal-to-noise ratio because it permits an increase in the input-signal level or an increase in the gain of the stages preceding the noise source without overloading the output stage. This means, for example, that the power supply for the output stage of a feedback amplifier requires less filtering as would normally be required.

(e) *Reduction of nonlinear distortion.* A feedback amplifier reduces the percentage of harmonic distortion in the output by a factor of $1/1 - \beta K$.

13-1.44 Disadvantages

The disadvantages of negative feedback are:

(a) Stability problems arise. These problems are analyzed and solved in the manner outlined in Section I of this publication.

(b) Over-all gain is less than in an open-loop amplifier with the same number of stages. That is,

$$K' = \frac{K}{1 - \beta K} \quad (13-44)$$

Since $|1 - \beta K| > 1$ for negative feedback, $K' < K$. To achieve a specified gain, a negative-feedback amplifier requires more stages of amplification than an open-loop amplifier.

13-1.45 PROBLEMS ENCOUNTERED IN USE OF ELECTRONIC AMPLIFIERS AS SERVO COMPONENTS

13-1.46 Reliability

The greatest single cause of failure in electronic servo amplifiers is vacuum-tube breakdown. In many modern designs, vacuum tubes have been replaced by transistors and magnetic amplifiers, but there are still cases where dynamic requirements or environmental conditions dictate the use of vacuum tubes. In these cases, the designer can do much toward achieving amplifier reliability. It is possible to design amplifiers that will operate for thousands of hours without failure by taking the following measures:

(a) *Care in selection of tube types.* There are now available higher-quality tube types than those originally developed for consumer use. The highest-quality tubes that are well-adapted to military applications are listed in MIL-STD-200. Tube types should be selected from this list whenever possible.

(b) *Derating of tubes.* Particular care must be taken to ensure that maximum tube ratings are not exceeded. These ratings are published in tube manufacturers' catalogs and in Military Specification MIL-E-1. Circuits should be designed for operation well below maximum tube ratings because tube life is appreciably increased when the tubes are thus derated. While there is no available analytical relationship giving tube life as a function of derating, ample experimental evidence shows that derating of tube characteristics leads to increased tube life. Maximum plate dissipation, maximum plate voltage, maximum cathode current, and maximum bulb

temperature are the major characteristics that should be derated.

(c) *Cooling of tubes.* Vacuum tubes should not be located near other heat-radiating parts and proper ventilation should always be provided. Where convection cooling cannot be made adequate, either by natural circulation or forced air flow (fans), conduction cooling can be provided by means of a metallic bond between the tube envelope and the chassis (e.g., by using a properly designed tube shield).

(d) *Interchangeability of tubes.* Circuits should be designed so that malfunction cannot be caused by variations in characteristics between tubes of the same type (as listed in MIL-E-1) or by reasonable variations in tube characteristics caused by aging. Negative feedback, for example, can do much to stabilize circuit performance in the presence of tube-characteristic variations. A properly designed circuit does not require a specially selected tube. When circuit design necessitates the selection of tubes with particular characteristics, unmanageable maintenance problems inevitably follow.

(e) *Protection from shock.* In mobile applications, tubes must be protected from mechanical shock. Shock mounts on the amplifier chassis are very helpful and, in addition, the amplifier should be located as far as possible from shock-producing members. Where mechanical shock is unavoidable, select tube types specifically designed to tolerate shock and use their acceleration ratings as a guide in this selection.

(f) *Filament-voltage regulation.* Filament voltages in excess of nominal ratings are particularly detrimental to tube reliability. If the source of heater voltage is known to be variable, it may be possible to select a center value that is lower than the tube rating, providing the resulting decrease in tube performance does not produce circuit malfunction. Otherwise, some means of heater-voltage control must be devised, or a more stable source provided.

(g) *Derating of passive circuit components.* Circuit components such as resistors and capacitors (called *passive* components) must be carefully chosen for highest quality. The effects of variations in environmental conditions on component electrical characteristics must be considered in the over-all amplifier design. Derating of manufacturers' specifications is helpful in extending the useful life of passive components as well as tubes. Many designers now derate the power-dissipation ratings of resistors and the maximum voltage ratings of capacitors (at normal ambient temperatures) by as much as 50 percent. Additional derating of these specifications is made for operation under unusual or extreme conditions. To include aging effects in resistors, some design groups also derate the resistance-tolerance specification and use a 1-percent resistor where a 5-percent tolerance is initially required, and a 5-percent resistor where 10- to 20-percent tolerance is required. While these measures lead to higher initial equipment cost, they often pay for themselves in reduced maintenance costs.

13-1.47 Construction

In the physical construction of an amplifier, the dominant considerations are: adherence to weight and space requirements; freedom from unwanted electrical coupling between components; proper ventilation to prevent excessive temperatures; freedom from mechanical vibration; and accessibility of components for maintenance.

13-1.48 Maintenance

Equipment down-time can be materially reduced by using plug-in subassemblies that can be quickly replaced by spares in case of failure. To reduce the required inventory of spare units, as many as possible should be identical. For example, it is frequently possible to use several identical *complete amplifiers* at different places in a system. The single design, in this case, must be capable of satisfying different requirements and may therefore require greater design effort. However, the resultant operational and maintenance convenience often makes this approach very desirable.

13-1.49 Quadrature Signals

In some applications, a servo amplifier is required to amplify a signal that contains two components, one component in time-phase with respect to a signal to which the output member of the servomechanism can respond and representing the loop error, and the other component 90° out of phase with the first. The second component, called the *quadrature* signal, serves no useful purpose. Its presence, however, causes the amplifier to saturate earlier than it otherwise would, thereby reducing the effective gain of the amplifier. For this reason, it is desirable to remove the quadrature-signal component. One method employs a keyed demodulator that is triggered at the time the instantaneous value of the quadrature voltage is zero. Since the output of a keyed demodulator will then be a function of the in-phase signal only, remaining constant between triggering points, it follows that the quadrature component is eliminated from the demodulator output.

13-1.50 Complete Amplifier

The design of a complete amplifier is not a rigidly determined procedure. The designer must take into consideration not only the amplifier specifications (load characteristics, input signal voltage and impedance, amplifier dynamics, space and weight restrictions), but also such characteristics of the over-all system as environmental conditions and available power supplies. However, the designer may very well be able to design several different amplifiers that satisfy the requirements. For example, after a push-pull power-amplifier stage has been designed, a freedom of choice often exists for selecting the means of obtaining the two grid signals. This choice could be a transformer with a center-tapped secondary, a plate- and cathode-loaded amplifier, or a differential amplifier with one grid grounded. In general, the objective is to achieve the required gain in the voltage-amplifier section with the least number of stages so that the tube complement is held to a minimum. On the other hand, it might be

desirable to incorporate negative feedback for gain stability. This decreases the open-loop gain and, as a result, the design might use more than the minimum possible number of tubes. Here, as in so much of systems design, trial-and-error procedures are used and even after a design is completed, building and testing a breadboard model nearly always results in a considerable number of design changes.

13-1.51 Details of a Typical Servo Amplifier

Figure 13-40 shows the circuit schematic of the main elements of a complete amplifier designed to drive a 2-phase motor. Tubes V4 and V5 form a push-pull amplifier. The control-grid signal for V4 is derived from voltage amplifier V3A, one of two triode sections in the same tube envelope. The control-grid signal for V5 is derived from V3B, the output of which is of the same magnitude as, but 180° out of phase with, the output of V3A. The input signal for V3B is part of the output of V3A, and hence V3B inverts that signal. The ratio $R_{22}/R_{21} + R_{22}$ and the gain of V3B are adjusted so that the output of V3B is exactly equal in magnitude to the output of V3A. V2B is another simple voltage amplifier and V2A is an electronic phase shifter comprising a phase splitter similar to the one discussed in connection with Fig. 13-21. By connecting R_9 and C_4 between the plate and cathode of V2A, the output voltage at the junction of R_9 and C_4 can be varied in phase from 7° to 170° without appreciable change in amplitude simply by varying R_9 . V1 is a summing amplifier for the two input signals. Each stage of the servo amplifier uses a cathode resistor to provide negative feedback within the stage and also to provide bias voltage, thereby eliminating the need for a negative bias-voltage supply. Series grid resistors R_{25} and R_{26} are used to suppress parasitic oscillations that frequently occur when the resistors are omitted. The capacitor across the output transformer compensates for the inductance of the motor; the capacitor value depends upon the particular motor used. Amplifier gain is controlled by varying R_{28} in the feedback loop. The gain factors for the

individual stages are indicated on the figure. Other specifications are:

Power Requirements:

For V4 and V5, Type 6V6 — plate current 80 ma at +300v, filament current 2.7 amp at 6.3 vac

For V4 and V5, Type 6L6 — plate current 110 ma at +300v, filament current 3.6 amp at 6.3 vac

Gain: 1000 min, 5000 max

Power output: 15 watts max

Load impedance: adjustable from 200 to 20,000 ohms through transformer taps

Noise level: less than 1 volt output with both inputs grounded

Internal phase shift: adjustable from 7° to 170°

Linearity: linear for input signals up to ± 50 mv

Response: flat within ± 1 db from 50 to 10,000 cps for resistive load of 1200 ohms; flat to ± 50 cps of the modulating frequency (3° phase shift at 50 cps), for motor loads, with carrier frequency constant; carrier frequencies of 60, 400, and 1000 cps can be used (for 60 cps carriers, the coupling capacitors should be increased to 0.05 microfarad).

13-1.52 THYRATRON AMPLIFIERS

Amplifiers employing thyratron tubes can be used in certain servomechanism applications to supply resistive, capacitive, or inductive loads. The thyratron is an efficient amplifying device that provides high power gain.

13-1.53 Description of Thyratron

A thyratron is a thermionic tube containing a plate, a cathode, and one or more grids. The thyratron envelope is filled with hydrogen, mercury vapor, or a noble gas such as xenon. Hydrogen-filled thyratrons are used only for high-voltage pulse work; they will not be considered in the following discussion.

Mercury-vapor-filled thyratrons must be kept between fairly narrow operating-temperature limits, usually between 40°C and 80°C, thus limiting the ambient temperatures in which these tubes can be used. However, nearly all

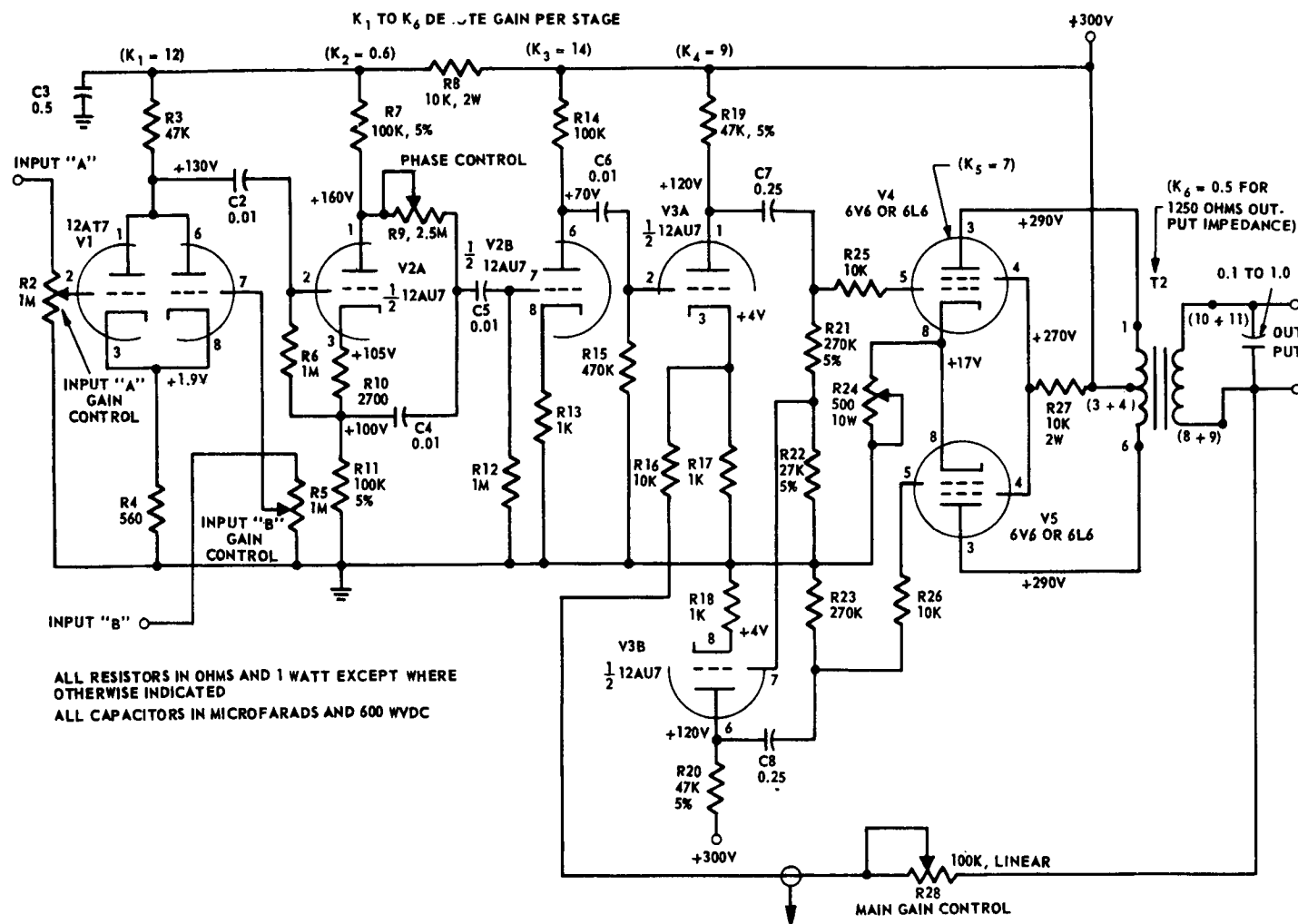


Fig. 13-40 A-c servo amplifier.

thyratrons filled with noble gas can be used in ambient temperatures ranging from -55°C to $+75^{\circ}\text{C}$, with some tubes having an even wider ambient-temperature operating range. The control power required by thyratrons is quite small, being generally about 100 milliwatts. Load-current ratings range from about 40 milliamperes to 18 amperes, and more. Peak-forward and peak-inverse voltage ratings range from about 350 to 2000 volts, with ratings from 750 to 1500 volts being the most common for the larger tubes. Plate-to-cathode drop is generally 8 to 15 volts.

13-1.54 Thyatron Characteristics

The thyatron acts essentially as a switch capable of conducting current in one direction only. As long as the voltage between control grid and cathode is more negative than the critical grid voltage, the thyatron does not conduct plate current. When the grid-to-cathode voltage is more positive than the critical grid voltage, the thyatron conducts plate current, provided that the plate is positive with respect to the cathode. However, the control grid loses all control over the plate current as soon as the thyatron begins to conduct, and current flows as long as the plate-to-cathode voltage is positive, even if the control-grid voltage is again made more negative than the critical grid voltage. Figure 13-41 shows typical control characteristics of a thyatron. The grid voltage at which conduction starts is called the *critical grid voltage* and is a function of plate voltage. The shaded area in the figure indicates a region of uncertainty where conduction may, or may not, start. This uncertainty region is the result of manufacturing variations, variation in tube temperature, and tube aging. Conduction in a thyatron is carried on by ionized gas. At the end of the conduction period, a certain time is required for the gas to be deionized. Usually, this deionization time is about 1 millisecond, thus placing a limit on the frequency of the alternating voltage that can be controlled by a thyatron. Nearly all thyratrons require a warm-up period, ranging from about 10 seconds to one minute, every time the equipment is started. This warm-up pe-

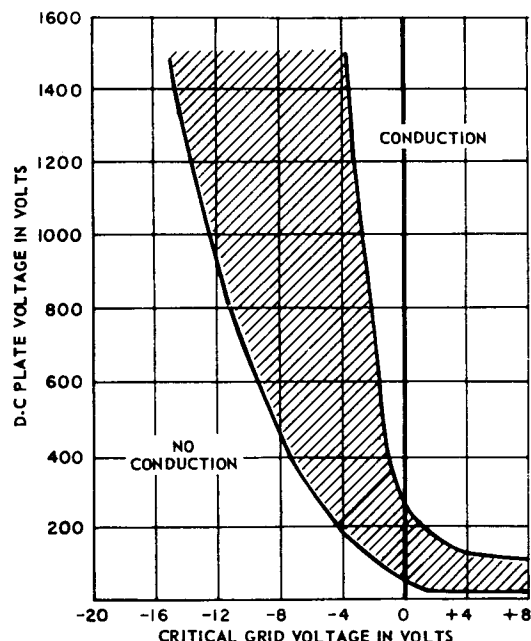


Fig. 13-41 Typical thyatron control characteristics.

riod prevents damage to the cathode by positive-ion bombardment.

Once a thyatron begins to conduct, the control grid has no further influence on the magnitude of the plate current, and grid control can only be re-established by removing the plate voltage or by reversing it. Removal of the plate voltage is nearly always a cumbersome procedure, but reversal can be accomplished by the simple expedient of using ac instead of dc to supply the plate voltage. For this reason, thyratrons are hardly ever supplied from a d-c source. Reversal of the alternating plate voltage interrupts the plate current during each cycle, thus re-establishing grid control for the next cycle.

13-1.55 Thyatron Amplifier with Resistive Load

Figure 13-42 shows a simple half-wave thyatron amplifier working into a purely

resistive load. The output current of this amplifier is a pulsating direct current, the magnitude of which depends upon the grid voltage. A common method of applying grid control is shown in this figure, where a constant sinusoidal voltage is phase-shifted by an R-C network so that it lags the plate voltage of the thyatron by 90° . This phase-shifted voltage, frequently called the *rider*, is added to a variable d-c control voltage E_c ; the sum of the two voltages is then applied between grid and cathode of the thyatron. A very small capacitor C_2 is connected directly between grid and cathode to prevent transient disturbances in the plate voltage from producing excessive voltages on the control grid through the plate-to-grid capacitance.

Figure 13-43 shows the waveforms of plate voltage and critical grid voltage of the thyatron as a function of time. The actual grid voltage produced by the circuit of Fig. 13-42 is shown for two values of the control voltage E_c , together with the resulting load voltage. If E_c is positive (Fig. 13-43A), the grid voltage exceeds the critical grid voltage early in the positive half-cycle of plate voltage. The thyatron thus "fires" early, permitting the plate voltage to be applied to the load over the

major portion of the positive half-cycle. The rectifying properties of the thyatron prevent a reversal of the load current; therefore, no negative voltage can be applied to the resistive load. If E_c is negative (Fig. 13-43B), the thyatron fires late in the positive half-cycle, and the plate voltage is applied to the load for only a small portion of the positive half-cycle.

13-1.56 Control of Load Voltage

With the grid circuit shown in Fig. 13-42, the load voltage can be controlled by varying the control voltage E_c . With resistive loads, the change in the d-c value of the load voltage is proportional to the change in the d-c control voltage. Thyatrons are often used in full-wave amplifiers, in which case a transformer or other means must be used to supply each tube with the appropriate rider. Occasionally, the a-c rider is purposely distorted by means of diodes, etc., to obtain more favorable control characteristics for inductive or capacitive loads.

The angle α at which the thyatron begins to conduct is often called the *firing angle*, or *angle of retard*. This firing angle can also be controlled by providing an a-c grid voltage,

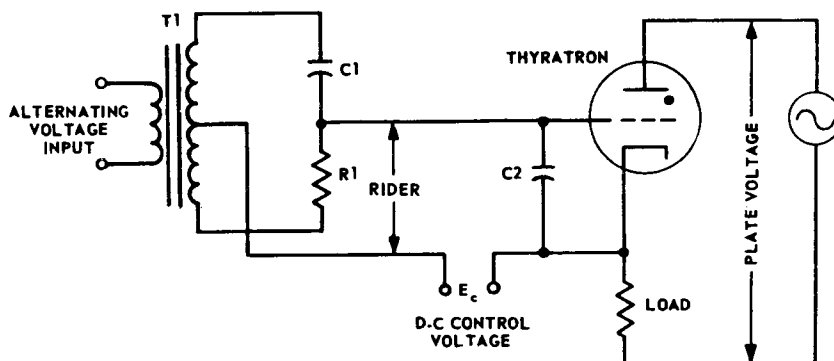


fig. 13-42 Half-wave thyatron amplifier.

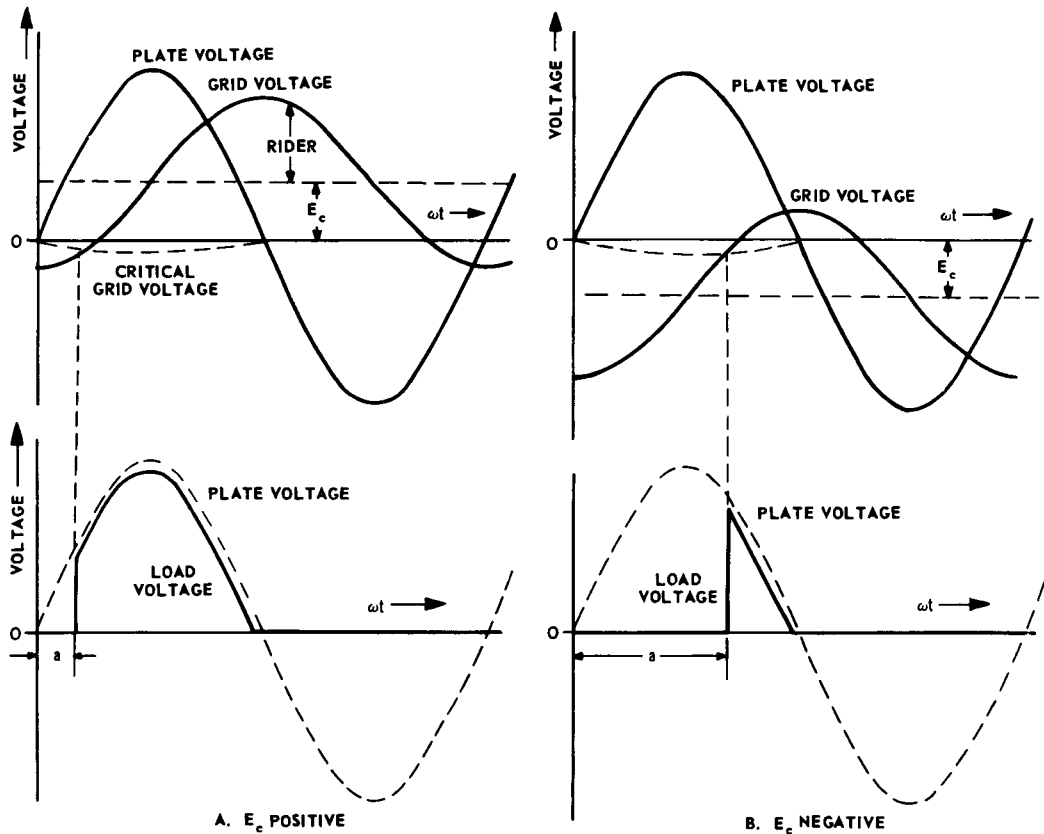


Fig. 13-43 Waveforms of circuit in Fig. 13-42.

the phase angle of which (with respect to the plate voltage) can be adjusted by external means. A simplified circuit of this type is shown in Fig. 13-44. The phase angle of the alternating grid voltage may be varied by means of various phase-shift circuits and devices, depending upon the application.

Control schemes furnishing a steep voltage pulse to fire the thyatron offer certain advantages in some applications. These pulses may be produced by electronic circuits or by magnetic devices such as pulse transformers. Another control scheme in this general category employs a magnetic amplifier to provide pulse signals, the phase angle of which (with

respect to the plate voltage) varies in response to the control current of the magnetic amplifier.

13-1.57 Thyatron-Amplifier Loads

Thyatrions can be used to supply loads consisting of resistive, inductive, or capacitive elements, either singly or in combination. If a d-c output is required, the thyatrions can be connected as single-phase or polyphase rectifiers of either the half-wave, full-wave, or bridge type. The back-to-back connection is used if an a-c load, such as the control winding of a servomotor, is to be supplied. Some of these circuits are shown in Fig. 13-45.

AMPLIFIERS USED IN CONTROLLERS

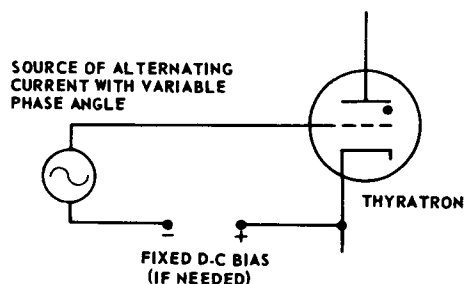
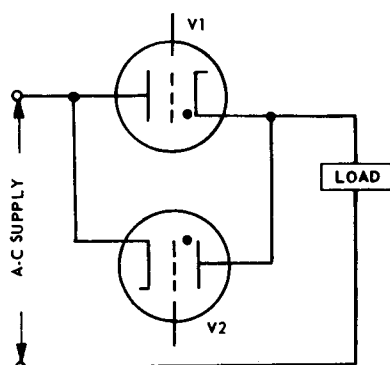


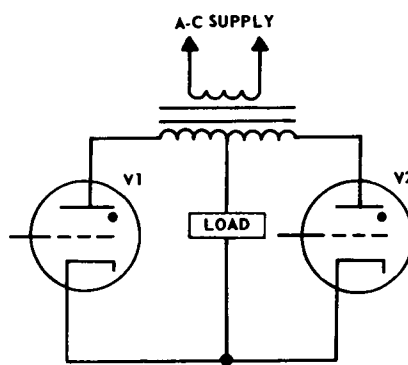
Fig. 13-44 Control of a thyatron by means of a phase-variable a-c signal.

13-1.58 Resistive loads. The load-voltage waveforms found in multitube circuits with resistive loads can be obtained by methods similar to those used to obtain the waveforms in Fig. 13-43.

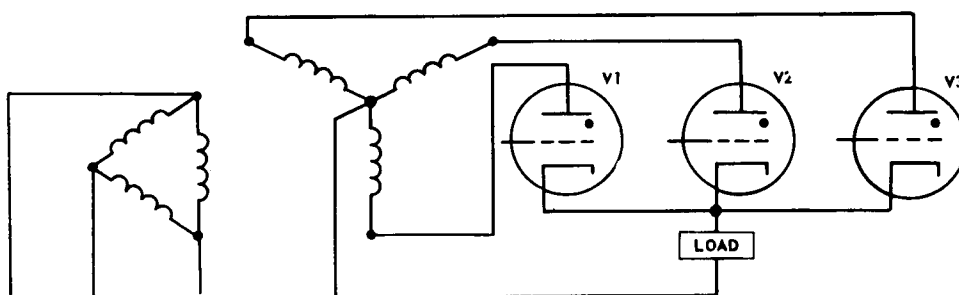
13-1.59 Inductive loads. If the load is inductive, the picture is somewhat more complicated. When the firing angle is very large, the load current is very small because, during any conduction period, the load inductance prevents build-up of the load current as shown in Fig. 13-46. As the firing angle is reduced, the load current increases slowly until the point is reached at which the load current supplied by one tube persists until the next tube



A. BACK-TO-BACK (SUPPLIES A-C TO LOAD)



B. SINGLE-PHASE FULL-WAVE (SUPPLIES D-C TO LOAD)



C. 3-PHASE HALF-WAVE (SUPPLIES D-C TO LOAD)

Fig. 13-45 Plate and load connections of typical thyatron amplifiers.

SERVOMECHANISMS

SECTION 4, POWER ELEMENTS AND SYSTEM DESIGN

**CONSISTING OF
CHAPTERS 14–20**

PREFACE

Section 4 of the handbook on Servomechanisms contains Chapters 14 through 20, which discuss servo power elements and system design. The significant features of servo output members are presented in Chapter 14 (Power Elements Used in Controllers) and in Chapter 15 (Mechanical Auxiliaries Used in Controllers). Chapter 16 is devoted to a discussion of the typical procedure that can be used as a guide when designing a servo system; this is followed by descriptions in Chapter 17 of two representative servo systems from existing Army equipments. Supplementary design information is given in Chapter 18 (Auxiliaries Associated with Servomechanisms), Chapter 19 (Constructional Techniques) and Chapter 20 (Supplementary Tables, Formulas, and Charts).

For information on other servomechanism components and on feedback control theory, see one of the following applicable sections of this handbook:

ORDP 20-136 Section 1 Theory (Chapters 1-10)

ORDP 20-137 Section 2 Measurement and Signal Converters (Chapters 11-12)

ORDP 20-138 Section 3 Amplification (Chapter 13)

An index for the material in all four sections is placed at the end of Section 4.

CHAPTER 14**POWER ELEMENTS USED IN CONTROLLERS*****14-1 INTRODUCTION**

The following types of power elements or motors used in servomechanisms are discussed in this chapter: direct-current and alternating-current electrical motors, hydraulic and pneumatic motors, and magnetic-particle clutches. A comparison between a-c and d-c

motors appears in Par. 14-2.1; problems of hydraulic motors, pneumatic motors, and magnetic clutches are listed in Pars. 14-4, 14-5, and 14-6, respectively. The choice of which output motor to use is discussed in Par. 16-3.4.

14-2 DIRECT-CURRENT MOTORS**14-2.1 USAGE OF D-C MOTORS**

Of the two classes of motors available for servo applications, direct-current motors, in comparison with alternating-current motors, have the following advantages:

- (a) Speed is easily controllable.
- (b) By varying field excitation, high power gains are possible for speed control.
- (c) High efficiency is obtained for motors larger than 100-watt rating.
- (d) Dependable rotary electric amplifiers to drive larger size motors are readily available.

Direct-current motors have the following disadvantages:

- (a) Commutator produces electrical noise.
- (b) Brushes wear out (conditions of cold, dry air appear to accelerate wear).

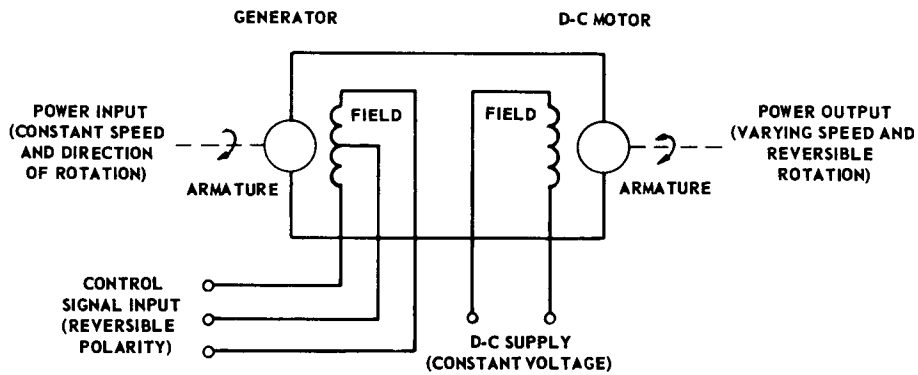
(c) Brushes have an undesirable electrical characteristic of acting as a nonlinear resistance (an almost constant voltage drop occurs across the brush; this voltage drop must be exceeded when starting the motor before a proportional input-output relationship can be obtained).

(d) Brushes add coulomb friction (for small motors, 10 to 20 watts, this becomes of consequence).

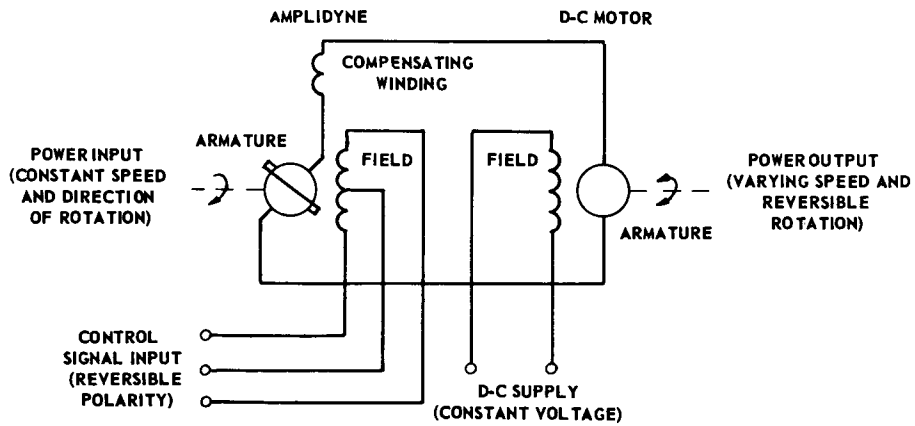
(e) The added complexity of a demodulator is required if an a-c signal is used in the error detector and amplifier.

The types of d-c motors used in servo applications are the series motor, the separately excited shunt motor, and the permanent-magnet motor. Series motors have high starting torque and poor speed regulation. Shunt motors have lower starting torque but better

*By P.E. Smith, Jr.



A. BASIC TYPE



B. AMPLIDYNE TYPE

Fig. 14-1 Rotary electric amplifier circuits for control of d-c motor.

speed regulation than series motors. Permanent-magnet motors are easy to drive since no power is used for the field. A limitation is that the field may be demagnetized if the motor is badly overloaded. Conventional d-c motors are unidirectional devices. The armature rotation of series and shunt motors can be reversed by reversing the current flow through either the armature or field but not both simultaneously. Permanent-magnet motors are reversed by reversing the direction of armature current; high torques on reversal are obtainable.

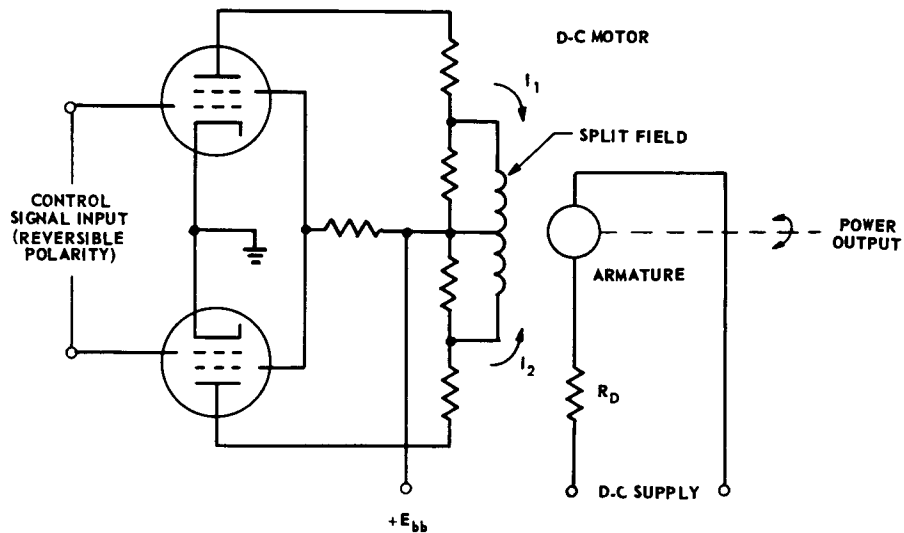
Many means are available for controlling the speed or armature position of d-c motors; they are

- (a) Rotary electric amplifiers
- (b) High-vacuum tube amplifiers
- (c) Gas-filled tube amplifiers
- (d) Magnetic amplifiers
- (e) Relay amplifiers

The generator, a basic rotary electric amplifier (Fig. 14-1A), operates at a constant speed and produces an output voltage that is proportional to the input signal applied to its field. The output voltage developed across the low impedance of the generator armature excites the d-c motor armature with comparatively low power losses. The amplifying action of the generator produces power gains up to 100. A variation of the basic rotary electric amplifier is the amplidyne (Fig. 14-1B). A short-circuited brush connection in the armature of the amplidyne makes it possible to obtain high power gains of 3000 to 10,000. By this means, it is feasible to control 5- to 10-hp motors from a source capable of delivering only 5 to 10 watts of input signal. The higher power gain of the amplidyne is achieved at the expense of a slower response. Rotary electric amplifiers (see Par. 13-4) are used when large motors are to be controlled; when drive motors less than 1/4-hp rating are to be controlled, this type of amplifier becomes impractical.

High-vacuum tube amplifiers (see Par. 13-1) are used to control d-c motors when the motor is small enough so that a rotary electric amplifier is impractical. The amplifier output may be used to excite either the armature or field of the motor. Armature control by high-vacuum tubes is limited to motors not larger than 1/20-hp rating because the entire motor output power is derived from the armature circuit. Field control (Fig. 14-2) may be used with motors whose ratings do not exceed 1/2 hp; this assumes an amplifier output of 20 watts and a motor field-to-armature power gain of about 25. To compensate for the normal characteristics of a field-excited motor to exhibit increased speed at decreased field excitation and reduced torque at high speed, a constant armature current is required. Partial compensation is obtained by using a current-limiting resistor in series with a high-voltage supply for the armature.⁽¹⁾ To prevent large induced voltages from being developed in the field windings when the field current is cut off for reversal, some protective device should be used to limit the voltage rise. For small motors, each field winding may be shunted by a fixed resistor.⁽²⁾

Gas-filled tube amplifiers (see Par. 13-1) can control motors about ten times larger than is feasible with high-vacuum tube amplifiers. Typical circuits using thyratrons with either a-c or d-c control signals are shown in Fig. 14-3.^(3,4,5) The split-field series motor control circuit shown in Fig. 14-3A uses no plate transformer. In Fig. 14-3B, the plate transformer provides circuit isolation from the a-c supply and a better match of a-c supply voltage to tube and motor characteristics. In the circuit shown in Fig. 14-3C, two gas-filled diodes allow a common cathode input circuit to be used without a plate transformer. When V_1 is conducting, armature current passes through it and diode V_4 . When V_2 is made conducting, current passes through it and diode V_3 . Resistors R_1 and R_2 maintain the common cathode connection at a fixed potential to ground during the time neither triode is conducting. For gas-filled tube amplifiers, Burnett⁽⁶⁾ states that the best system response may be obtained by the use



$$R_D \geq \frac{V_o}{I_R} - R_a$$

WHERE

R_D = CURRENT-LIMITING RESISTANCE

I_R = RATED ARMATURE CURRENT

V_o = D-C SUPPLY VOLTAGE

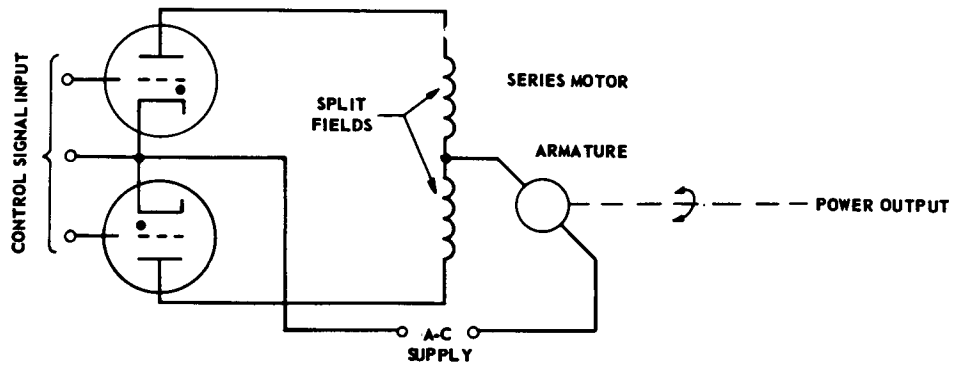
R_a = ARMATURE RESISTANCE

$I_1 > I_2$ LEADS TO CW ROTATION

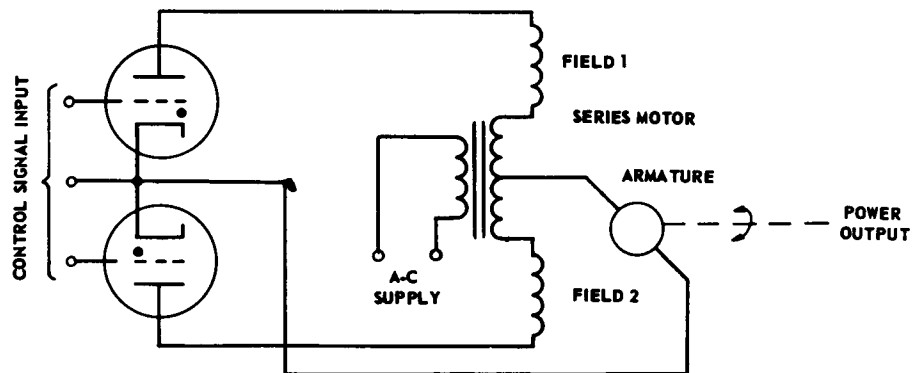
$I_2 > I_1$ LEADS TO CCW ROTATION

Fig. 14-2 High-vacuum tube circuit for control of d-c motor.

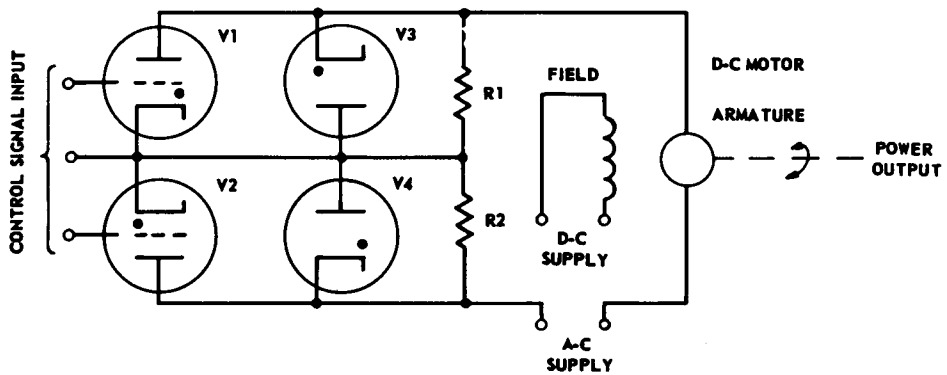
POWER ELEMENTS USED IN CONTROLLERS



A. SPLIT-FIELD SERIES MOTOR (NO PLATE TRANSFORMER)



B. SPLIT-FIELD SERIES MOTOR (WITH PLATE TRANSFORMER)



C. SHUNT MOTOR (NO PLATE TRANSFORMER)

Fig. 14-3 Gas-filled tube circuits for control of d-c motors.

of armature control powered from a multi-phase supply. He also states that, with 3-phase power, a 10-hp velocity servo with a frequency response flat to 15 cps is possible; position control with a flat response to 10 cps with the same motor is claimed.

Magnetic amplifiers (see Par. 13-3) provide more reliable control of d-c motors than do vacuum-tube or relay amplifiers. Figure 14-4 shows a typical magnetic amplifier with two cores for controlling the position of a series motor with split-field windings.⁽⁷⁾ By reversing the polarity of the input signal, either direction of motor rotation can be obtained. If a separately excited motor with constant-voltage field supply were used, excessive circulating current would result when the motor operates in one direction.⁽⁸⁾ Magnetic amplifiers can be used to control the speed of d-c motors by field control or armature control.⁽⁹⁾ One circuit (Fig. 14-5) for unidirectional speed control uses a large inductance in series with the armature circuit to

obtain sloping speed-torque characteristics very similar to the uncontrolled motor but with greater slope.⁽¹⁰⁾ If compensation proportional to the armature current is added, nearly flat speed-torque characteristics can be achieved.

Relay amplifiers (see Par. 13-5) are used to control d-c motors when size and weight are primary considerations and reliability is secondary. A single-pole, double-throw polarized relay (Fig. 14-6A) provides a simple means for controlling a series motor with split-field windings. No motor damping is provided when the relay is not energized (contacts open). If motor damping is desired, two single-pole, double-throw relays which are alternately energized may be used (Fig. 14-6B). To minimize relay contact current, field control of a shunt motor can be used. In this case, the armature should be powered from a constant current source. The protection of relay contacts is discussed in Par. 13-5.

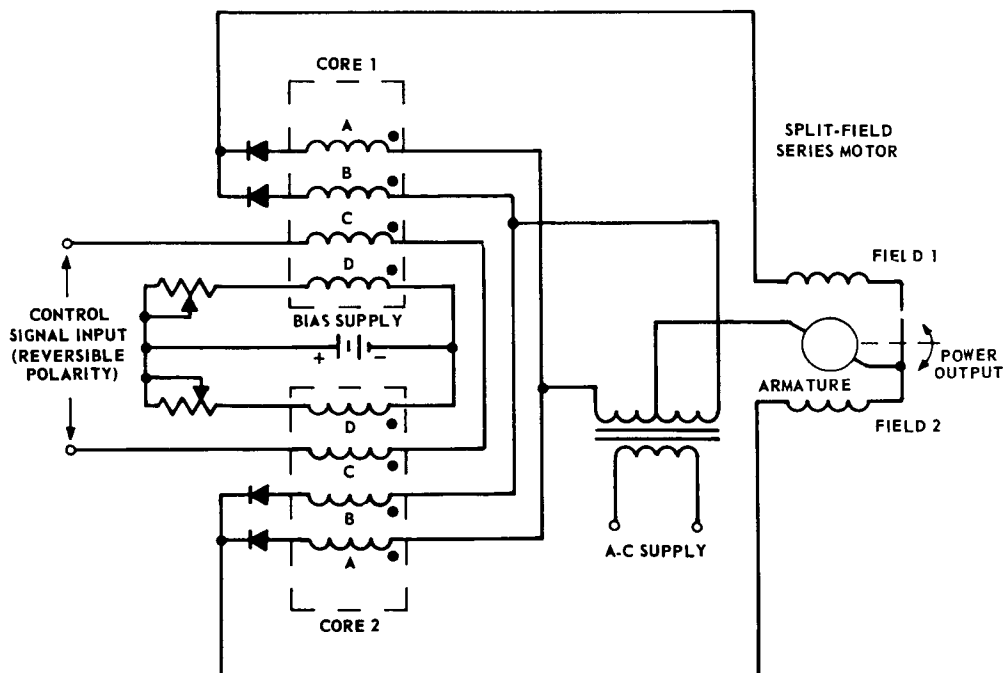


Fig 14-4 Magnetic amplifier circuit for control of series motor.

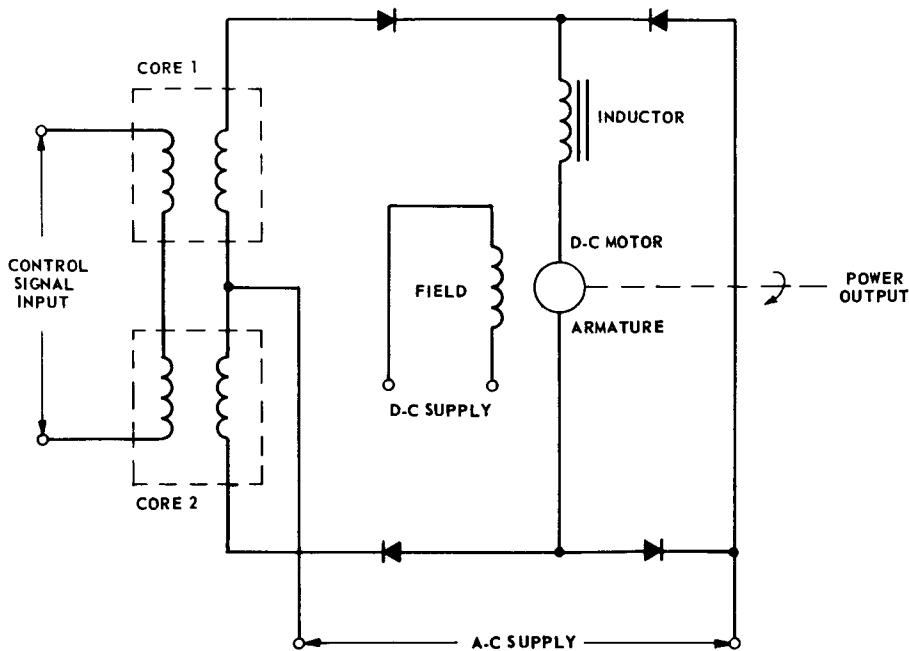


Fig. 14-5 Magnetic amplifier circuit for unidirectional control of shunt motor.

14-2.2 STATIC CHARACTERISTICS OF D-C MOTORS

In the general case for all types of d-c motor control, the following equations express the steady-state behavior of the motor:

$$V_t = I_a R_a + K I_f \omega \quad (14-1)$$

$$T = K I_f I_a \quad (14-2)$$

where

V_t = armature terminal voltage, in volts

I_a = armature current, in amperes

I_f = field current, in amperes

R_a = armature resistance, in ohms

K = a constant

T = torque produced by motor, in newton-meters

ω = speed of armature-shaft rotation, in radians/sec ($\omega = 2\pi \times \text{rpm}/60$)

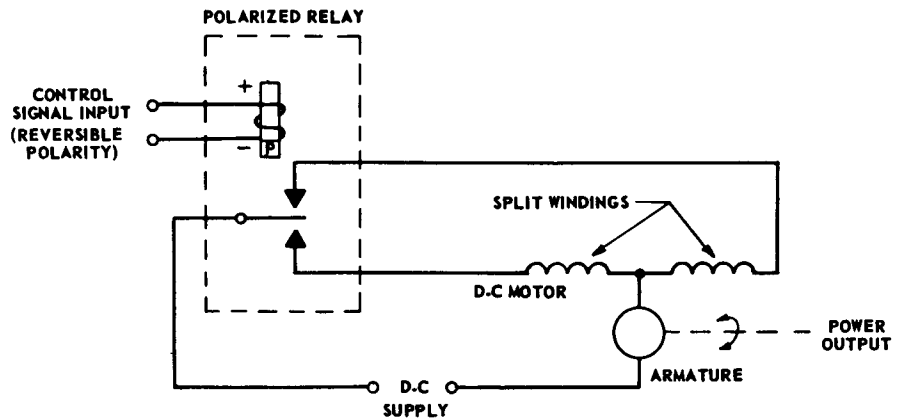
The above general equations may be modified for the following particular types of motor control:

- (a) Armature control with constant field current
- (b) Field control with constant armature current
- (c) Field control with constant armature voltage
- (d) Series motor control

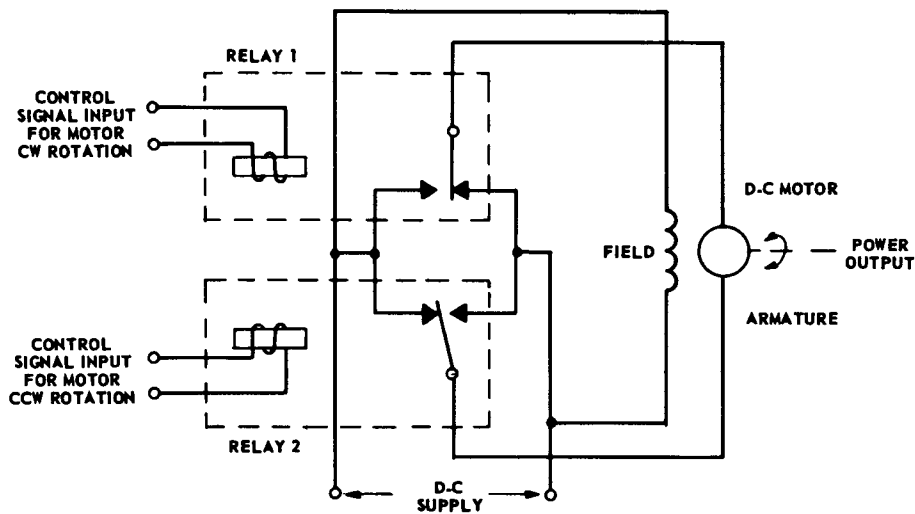
14-2.3 Armature Control with Constant Field Current

Since the field current is constant in this type of control, the term $K I_f$ may be combined in a new term K_1 , and substituted into Eqs. (14-1) and (14-2). Then Eq. (14-2) is substituted into Eq. (14-1) which is solved for ω to obtain

$$\omega = \frac{V_t}{K_1} - T \frac{R_a}{K_1^2} \quad (14-3)$$



A. SERIES MOTOR CONTROL (NO DAMPING)



B. SHUNT MOTOR CONTROL (WITH DAMPING)

Fig. 14-6 Relay amplifier circuits for control of d-c motors.

Solving Eq. (14-3) for T

$$T = \frac{K_1}{R_a} V_t - \frac{K_1^2}{R_a} \omega \quad (14-4)$$

The stall torque T_M which occurs at the maximum armature terminal voltage V_{tM} is found from Eq. (14-4) when $T = T_M$, $V_t = V_{tM}$, and $\omega = 0$.

$$T_M = \frac{K_1}{R_a} V_{tM} \quad (14-5)$$

Solving Eq. (14-5) for $\frac{K_1}{R_a}$

$$\frac{K_1}{R_a} = \frac{T_M}{V_{tM}} \quad (14-6)$$

Substituting Eq. (14-6) into Eq. (14-4)

$$T = \frac{T_M}{V_{tM}} V_t - \frac{T_M}{V_{tM}} K_1 \omega \quad (14-7)$$

The no-load speed ω_M at the maximum armature terminal voltage V_{tM} is found from Eq. (14-3) when $\omega = \omega_M$, $V_t = V_{tM}$, and $T = 0$.

$$\omega_M = \frac{V_{tM}}{K_1} \quad (14-8)$$

Solving Eq. (14-8) for $\frac{K_1}{V_{tM}}$

$$\frac{K_1}{V_{tM}} = \frac{1}{\omega_M} \quad (14-9)$$

Substituting Eq. (14-9) into Eq. (14-7)

$$T = \frac{T_M V_t}{V_{tM}} - \frac{T_M \omega}{\omega_M} \quad (14-10)$$

Dividing Eq. (14-10) by T_M results in

$$\frac{T}{T_M} = \frac{V_t}{V_{tM}} - \frac{\omega}{\omega_M} \quad (14-11)$$

A plot of T/T_M versus ω/ω_M characteristics with V_t/V_{tM} as a parameter is shown in Fig. 14-7. The curves of this figure are nondimensionalized and based on Eq. (14-11).

If the manufacturer's data for a particular motor are known, the curves can be dimensionalized for that motor. For example, a motor has an armature resistance of 5 ohms and a no-load speed of 3000 rpm at a maximum armature terminal voltage of 90 volts.

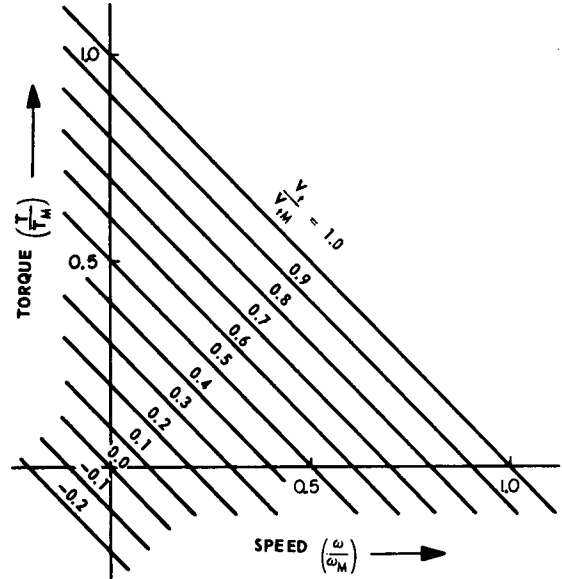


Fig. 14-7 Static torque-speed characteristics of armature-controlled d-c motor.

To dimensionalize a plot of T/T_M versus ω/ω_M characteristics, find the value of no-load speed ω_M in radians/sec and the value of stall torque T_M .

From Eq. (14-1), ω is defined as $2\pi \times$ rpm/60. At no-load speed, $\omega = \omega_M$ which is

$$\begin{aligned} \omega_M &= \frac{2\pi \times 3000}{60} \\ &= 314 \text{ radians/sec} \end{aligned}$$

At the point $\omega/\omega_M = 1$ on Fig. 14-7, the no-load speed ω_M is 314 radians/sec. All other points on the ω/ω_M axis are dimensioned proportionally.

Substituting 314 for ω_M and 90 for V_{tM} in Eq. (14-9) and solving for K_1 gives

$$\begin{aligned} K_1 &= \frac{V_{tM}}{\omega_M} \\ &= \frac{90}{314} \\ &= 0.286 \end{aligned}$$

From Eq. (14-5), the stall torque is

$$\begin{aligned} T_M &= \frac{K_1}{R_a} V_{tM} \\ &= \frac{0.286}{5} \times 90 \\ &= 5.15 \text{ newton-meters} \end{aligned} \quad (14-5)$$

At the point $T/T_M = 1$ on Fig. 14-7, the stall torque $T_M = T = 5.15$ newton-meters. All other points on the T/T_M axis are dimensioned proportionally. A value of 90 volts is assigned to curve $V_t/V_{tM} = 1$ (maximum armature terminal voltage), and the other curves are dimensioned proportionally.

14-2.4 Field Control with Constant Armature Current

The armature current must be maintained constant for linear operation of field-controlled motors. The torque is given by Eq. (14-2) which applies directly in this case.

$$T = KI_f I_a \quad (14-2)$$

With the maximum field current I_{fM} applied, the maximum torque T_M which results is

$$T_M = KI_{fM} I_a \quad (14-12)$$

Dividing Eq. (14-2) by Eq. (14-12) results in

$$\frac{T}{T_M} = \frac{I_f}{I_{fM}} \quad (14-13)$$

Figure 14-8 is a plot of T/T_M versus ω/ω_M characteristics with I_f/I_{fM} as a parameter. Note that the figure and Eq. (14-13) are nondimensionalized and that the torque is independent of the motor speed.

14-2.5 Field Control with Constant Armature Voltage

Equations (14-1) and (14-2) express the steady-state behavior of this type of control. The nondimensionalized equation is*

$$\frac{T}{T_M} = \left(\frac{I_f}{I_{fM}} \right) \left(1 - \frac{I_f}{I_{fM}} \times \frac{\omega}{\omega_B} \right) \quad (14-14)$$

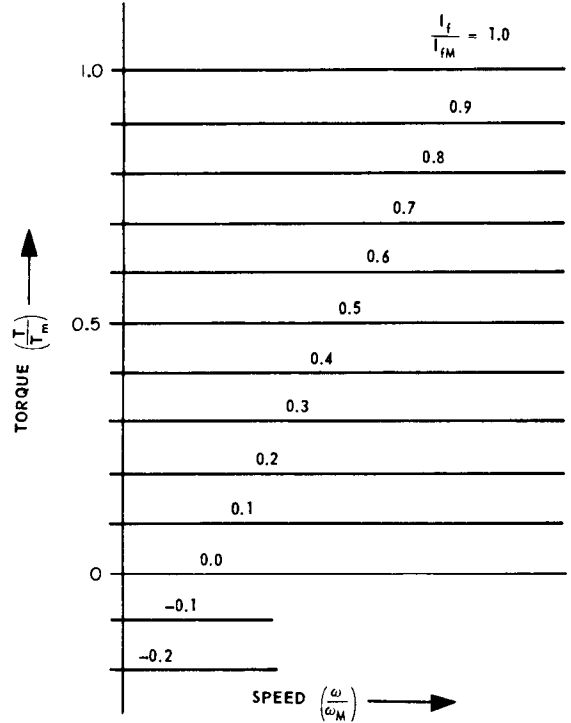


Fig. 14-8 Static torque-speed characteristics of field-controlled d-c motor, constant armature current.

where

I_{fM} = maximum field current

T_m = stall torque at $I_f = I_{fM}$

ω_B = no-load (zero-torque) speed for $I_f = I_{fM}$

A plot of Eq. (14-14) is shown in Fig. 14-9.

14-2.6 Series Motor Control

Since the armature current is the same as the field current in this type of control, the term I_f can be substituted into Eqs. (14-1) and (14-2) to obtain

$$V_t = I_a R_a + KI_a \omega \quad (14-15)$$

$$T = KI_a^2 \quad (14-16)$$

*See Appendix on page 14-24.

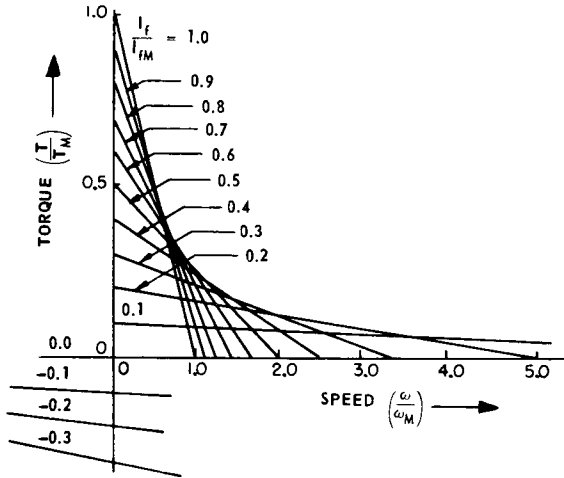


Fig. 14-9 Static torque-speed characteristics of field-controlled d-c motor, constant armature voltage.

Solving Eq. (14-16) for I_a and substituting into Eq. (14-15) gives

$$V_t = \left(\frac{T}{K} \right)^{1/2} \times \left(R_a + K \omega \right) \quad (14-17)$$

By defining $\omega_o = R_a/K$ and T_M as the stall torque at V_{tM} , the maximum armature terminal voltage, it can be shown that

$$\frac{T}{T_M} = \frac{\left(\frac{V_t}{V_{tM}} \right)^2}{\left(1 + \frac{\omega}{\omega_o} \right)^2} \quad (14-18)$$

A plot of T/T_M versus ω/ω_o characteristics with V_t/V_{tM} as a parameter is shown in Fig. 14-10. The curves in this figure are non-dimensionalized and based on Eq. (14-18).

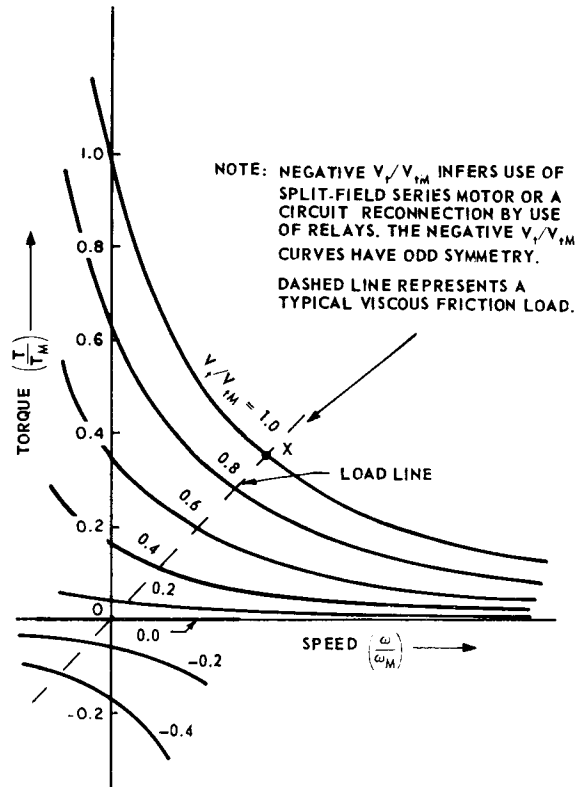


Fig. 14-10 Static torque-speed characteristics of series motor.

14-2.7 DYNAMIC CHARACTERISTICS OF D-C MOTORS

The dynamic characteristics of d-c motors involve additional parameters besides those that describe the static characteristics. The effect of the armature or field inductance and the mechanical inertia and load must be considered. The dynamic characteristics of the following types of motor control are discussed:

- Armature control with constant field current
- Field control with constant armature current
- Field control with constant armature voltage
- Series motor control
- Rotary electric amplifier control
- Electronic amplifier control

14-2.8 Armature Control with Constant Field Current

The dynamic characteristics for the armature-controlled motor are expressed by the equations

$$V_t = I_a R_a + K I_f \omega + L_a s I_a \quad (14-19)$$

$$T = T_{Lt} + J s \omega \quad (14-20)$$

where

s = the complex frequency operator

L_a = armature inductance, in henries

J = inertia of motor, gear train, and load referred to the motor shaft, in kilogram-meters²

T_{Lt} = load, bearing, and motor windage torque referred to the motor, in newton-meters

If the load, bearing, and motor windage torque has a component which can be approximated as proportional to the speed ω , the load torque may be expressed as

$$T_{Lt} = T_L + B\omega \quad (14-21)$$

where

$B\omega$ = viscous friction component of load torque

T_L = all effects of load torque except J and $B\omega$

Equations (14-19), (14-20), and (14-21) are combined and expressed in block diagram form in Fig. 14-11. Block diagram formulation is explained in Par. 3-5.

14-2.9 Field Control with Constant Armature Current

The following equations describe the dynamic behavior of a field-controlled motor, assuming an ideal constant-current source for the armature:

$$V_f = I_f R_f + L_f s I_f \quad (14-22)$$

$$T = T_L + B\omega + J s \omega \quad (14-23)$$

where

V_f = field terminal voltage, in volts

R_f = field resistance, in ohms

L_f = field inductance, in henries

Equations (14-22) and (14-23) are combined and expressed in block diagram form in Fig. 14-12.

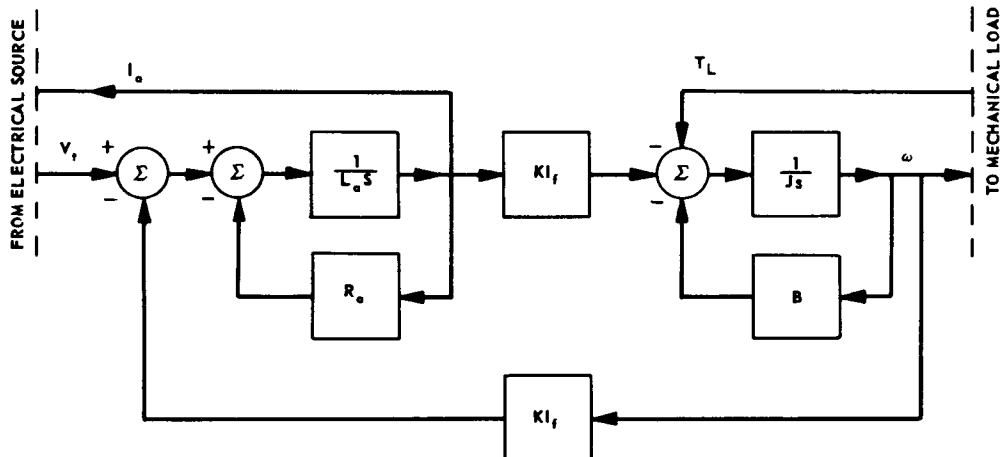


Fig. 14-11 Block diagram for armature-controlled d-c motor.

14-2.10 Field Control with Constant Armature Voltage

The dynamic behavior for the field-controlled motor with constant armature voltage is expressed by the equations

$$V_t = I_a R_a + K I_f \omega + L \frac{dI_a}{dt} \quad (14-24)$$

$$T = K I_f I_a \quad (14-2)$$

$$T = T_L + B\omega + J \frac{d\omega}{dt} \quad (14-25)$$

The terms $K I_f \omega$ and $K I_f I_a$ contain the products of two variables, and introduce a non-linearity. If the back-emf term $K I_f \omega$ is denoted as V_b , the variation in the quantities V_b and T can be written as

$$\Delta V_b = \frac{\partial V_b}{\partial I_f} \Delta I_f + \frac{\partial V_b}{\partial \omega} \Delta \omega \quad (14-26)$$

$$\Delta T = \frac{\partial T}{\partial I_f} \Delta I_f + \frac{\partial T}{\partial I_a} \Delta I_a \quad (14-27)$$

Add the average value V_{bo} to the variation ΔV_b ; and add the average value T_o to the variation ΔT . Then

$$V_b = V_{bo} + \Delta V_b = V_{bo} + \frac{\partial V_b}{\partial I_f} \Delta I_f + \frac{\partial V_b}{\partial \omega} \Delta \omega \quad (14-28)$$

$$T = T_o + \Delta T = T_o + \frac{\partial T}{\partial I_f} \Delta I_f + \frac{\partial T}{\partial I_a} \Delta I_a \quad (14-29)$$

The partial derivatives become

$$\left. \begin{aligned} \frac{\partial V_b}{\partial I_f} &= K\omega \\ \frac{\partial V_b}{\partial \omega} &= K I_f \\ \frac{\partial T}{\partial I_f} &= K I_a \\ \frac{\partial T}{\partial I_a} &= K I_f \end{aligned} \right\} \quad (14-30)$$

The average values of the various variables are: terminal voltage V_{to} ; armature current I_{ao} ; speed ω_o ; torque T_o ; and field current I_{fo} . In terms of average values only, Eqs. (14-24), (14-2), and (14-25) become

$$V_{to} = I_{ao} R_a + K I_{fo} \omega_o \quad (14-31)$$

$$T_o = K I_{fo} I_{ao} \quad (14-32)$$

$$T_o = T_{Lo} + B\omega_o \quad (14-33)$$

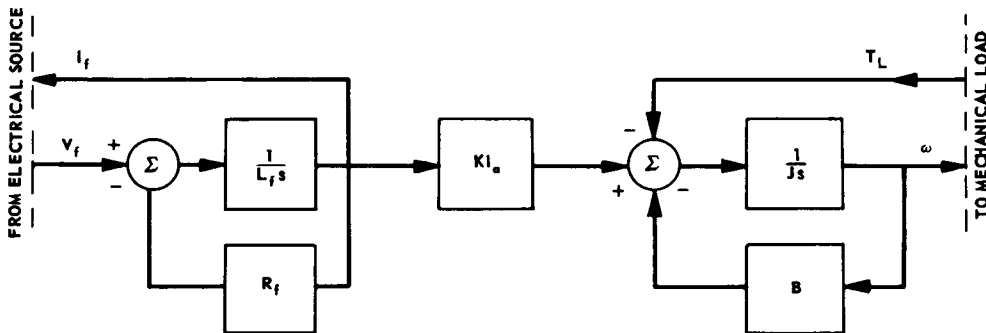


Fig. 14-12 Block diagram for field-controlled d-c motor, constant armature current.

The average field voltage is the product of the average field current and the field resistance; i.e., $V_{fo} = I_{fo}R_f$. For a position servo, ω_o can be considered as zero. In this case

$$V_{to} = I_{ao} R_a \quad (14-34)$$

$$T_o = KI_{fo}I_{ao} = T_{Lo} \quad (14-35)$$

Insert the expressions

$$I_a = I_{ao} + \Delta I_a$$

$$KI_f\omega = V_b = V_{bo} + \Delta V_b$$

$$\omega = \omega_o + \Delta\omega$$

$$I_f = I_{fo} + \Delta I_f$$

etc. into Eqs. (14-24), (14-2), and (14-25) after making use of Eqs. (14-28) through (14-30) to obtain

$$V_t = R_a(I_{ao} + \Delta I_a) + L \frac{d\Delta I_a}{dt} + KI_{fo}\omega_o + K(\omega_o + \Delta\omega)\Delta I_f + K(I_{fo} + \Delta I_f)\Delta\omega \quad (14-36)$$

$$T = KI_{fo}I_{ao} + K(I_{ao} + \Delta I_a)\Delta I_f + K(I_{fo} + \Delta I_f)\Delta I_a \quad (14-37)$$

$$T = J \frac{d(\Delta\omega)}{dt} + B\omega_o + B\Delta\omega + T_{Lo} + \Delta T_L \quad (14-38)$$

Substitute the average values given by Eqs. (14-34) and (14-35) into Eqs. (14-36) through (14-38) to obtain

$$V_t = I_{ao}R_a + \Delta I_a R_a + L \frac{d(\Delta I_a)}{dt} + 2K\Delta\omega \Delta I_f + KI_{fo} \Delta\omega \quad (14-39)$$

$$T_o = KI_{fo}I_{ao} + KI_{ao} \Delta I_f + 2K\Delta I_a \Delta I_f + KI_{fo} \Delta I_a = J \frac{d(\Delta\omega)}{dt} + B\omega_o t + B\Delta\omega + T_{Lo} + \Delta T_L \quad (14-40)$$

Subtract $V_{to} = I_{ao}R_a$ from Eq. (14-39) and $KI_{fo}I_{ao} = T_{Lo}$ from Eq. (14-40) and let all second-order differentials be zero (recall that $\omega_o = 0$) to obtain

$$\Delta V_t = 0 = \Delta I_a R_a + L \frac{d(\Delta I_a)}{dt} + KI_{fo} \Delta\omega \quad (14-41)$$

$$\Delta T = KI_{a0} \Delta I_f + KI_{f0} \Delta I_a$$

$$= J \frac{d(\Delta\omega)}{dt} + B\Delta\omega + \Delta T_L \quad (14-42)$$

Solve Eq. (14-41) for ΔI_a and use the Laplace notation to obtain

$$\Delta I_a = - \frac{KI_{f0} \Delta\omega}{R_a(1 + \tau s)} \quad (14-43)$$

where

$$\tau = L/R_a$$

Substitute Eq. (14-43) into Eq. (14-42).

$$\begin{aligned} \Delta T &= KI_{a0} \Delta I_f - \frac{(KI_{f0})^2 \Delta\omega}{R_a(1 + \tau s)} \\ &= (Js + B)\Delta\omega + \Delta T_L \end{aligned} \quad (14-44)$$

From Eqs. (14-34) and (14-35)

$$\frac{(KI_{f0})^2}{R_a} = \frac{T_{L0}^2}{I_{a0}^2 R_a} = \frac{T_{L0}^2 R_a}{V_{t0}^2} \quad (14-45)$$

Substitute Eq. (14-45) into Eq. (14-44).

$$\begin{aligned} \Delta T &= KI_{a0} \Delta I_f - \left(\frac{T_{L0}^2 R_a}{V_{t0}^2} \right) \frac{\Delta\omega}{R_a(1 + \tau s)} \\ &= (Js + B)\Delta\omega + \Delta T_L \end{aligned} \quad (14-46)$$

Equation (14-46) is shown in block diagram form in Fig. 14-13. Note that in Fig. 14-13 all variables are increments; the Δ notation is dropped for convenience. If $T_{L0} = 0$, Fig. 14-13 reduces to that of Fig. 14-12.

14-2.11 Series Motor Control

The dynamic characteristics for the series-connected motor are expressed by the equations

$$V_t = I_a R_a + KI_a \omega + L_a s I_a \quad (14-47)$$

$$T = T_L + B\omega + Js\omega \quad (14-23)$$

The runaway speed of the series motor depends upon the load; for example, point X

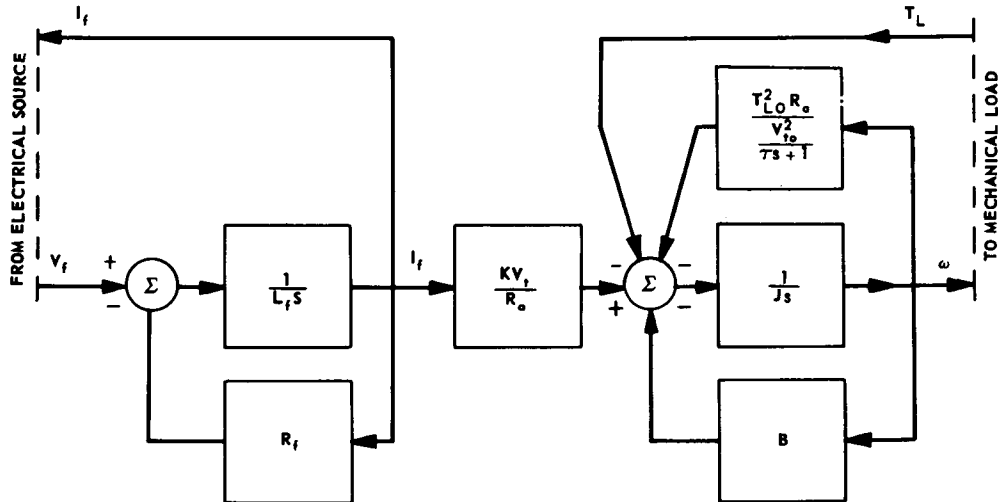


Fig. 14-13 Block diagram for field-controlled d-c motor, constant armature terminal voltage, incremental behavior.

shown in Fig. 14-10 indicates the runaway speed for a certain viscous load. By neglecting the armature inductance L_a and by assuming negligible dynamics throughout the rest of the system, it is possible to make a phase-plane solution of Eqs. (14-47) and (14-23). If the describing function method of solution to the problem is contemplated, computer solutions of Eqs. (14-47) and (14-23) for $\omega(t)$ will allow determination of an approximate transfer function. See Ch. 10 for an explanation of the describing function method and phase-plane solution of nonlinear problems.

14-2.12 Rotary Electric Amplifier Control

To compute the dynamic response of a rotary electric amplifier, its transfer function must take the source into account. For the

basic rotary electric amplifier (Fig. 14-1A), the block diagram shown in Fig. 14-11 can be used to represent its operation if the generator armature inductance and resistance are combined with that of the driven d-c motor. Also, the generated voltage V_o is assumed to equal the motor terminal voltage V_t . By making term q of Eq. (13-72) equal to k and equal to unity, this equation may be rewritten as

$$V_o = \frac{\alpha_1 V_1(s)}{1 + \tau_1 s} = \frac{\alpha_f V_f(s)}{1 + \tau_f s} \quad (14-48)$$

where the terms are as defined in Par. 13-4.

The transfer function relating the motor speed ω to the generator field voltage V_f and the load torque T_L is given by the following equation:

$$\omega(s) = \frac{\left[\frac{K_o}{1 + K_o} \right] \left[\frac{1}{KI_f} \right] \left[\frac{\alpha_f V_f(s)}{1 + \tau_f s} \right] - \left[\frac{T_2 s + 1}{B(1 + K_o)} \right] \left[T_L(s) \right]}{(T_a s + 1)(T_b s + 1)} \quad (14-49)$$

where

$$K_o = \frac{(KI_f)^2}{R_T B_T}$$

R_T = total armature resistance of generator and driven motor, in ohms

B_T = total viscous damping referred to motor shaft, in newton-meter-sec

$$T_a = \frac{T_1 + T_2 + \sqrt{(T_1 + T_2)^2 - 4(1 + K_o)(T_1 T_2)}}{1 + K_o}$$

$$T_b = \frac{T_1 + T_2 - \sqrt{(T_1 + T_2)^2 - 4(1 + K_o)(T_1 T_2)}}{1 + K_o}$$

POWER ELEMENTS USED IN CONTROLLERS

$$T_1 = \frac{J_T}{B_T}$$

$$T_2 = \frac{L_T}{R_T}$$

J_T = total inertia referred to motor shaft, in kilogram-meters²

L_T = total armature inductance of generator and driven motor, in henries

KI_f = stall torque to armature current ratio of motor

For larger motors, it may occur that Eq. (14-49) yields complex values of T_a and T_b denoted as $(a + jb)$ and $(a - jb)$. In this case, Eq. (14-49) can be rewritten with a new denominator resulting in

$$\omega(s) = \frac{\left[\frac{K_o}{1 + K_o} \right] \left[\frac{1}{KI_f} \right] \left[\frac{\alpha_f V_f(s)}{1 + \tau_f s} \right] - \left[\frac{T_2 s + 1}{B(1 + K_o)} \right] \left[T_L s \right]}{\left(\frac{s}{\omega_n} \right)^2 + 2 \frac{\zeta s}{\omega_n} + 1} \quad (14-50)$$

where

$$\omega_n = \frac{1}{\sqrt{a^2 + b^2}}$$

$$\zeta = a\omega_n$$

For the case of the amplidyne type of rotary electric amplifier (Fig. 14-1B), the expression for the generated voltage V_o given by Eq. (14-48) must be modified. Refer to Par. 13-4.

14-2.13 Electronic Amplifier Control

Two cases of electronic amplifier control are considered — one for the armature-driven motor, and the other for the field-driven motor. When an amplifier is used to drive an armature, and the armature is in the plate circuit, the amplifier can be represented by an equivalent voltage source ($V_o = \mu E_g$) in series with an internal resistance R_s . If the internal resistance is added to the armature resistance, Eq. (14-49) will apply by letting τ_f be zero, V_f be E_g , and α_f be μ .

When an amplifier is used to drive a motor field (constant armature current), the transfer function that determines the dynamic response is found by the aid of Fig. 14-12 to be

$$\omega(s) = \frac{\frac{\mu E_g(s)}{R_T(T_c s + 1)} - \frac{T_L(s)}{B}}{T_a s + 1} \quad (14-51)$$

where

$$T_c = \frac{L_f}{R_f + R_s}$$

R_s = amplifier internal or source resistance, in ohms

$$R_T = R_f + R_s$$

$$T_a = \frac{J}{B}$$

$$\mu = \text{amplifier gain} = \frac{\Delta V_f}{\Delta E_g}$$

14-2.14 MEASUREMENT OF D-C MOTOR PARAMETERS⁽¹¹⁾

The parameters of the d-c motor that are of interest in servomechanism applications are

R_a = armature resistance

L_a = armature inductance

R_f = field resistance

L_f = field inductance

K = motor constant

J = armature inertia

B = viscous damping

These parameters may include contributions from the source or from the load attached to the motor. In the measurements described in the paragraphs that follow, only the contributions of the motor itself to the parameters are considered.

14-2.15 Armature Resistance

If the manufacturer's data for armature resistance R_a are not available, it may be determined by the application of Ohm's Law, $R = E/I$. While keeping the armature blocked to prevent rotation, the applied armature voltage necessary to maintain rated armature current is measured. In the absence of armature rotation, there is no back-emf to oppose the applied armature voltage; therefore, a much lower applied voltage than rated voltage will produce rated current. The applied voltage at various angular armature positions should be measured and averaged to calculate the resistance R_a . The applied voltage should be connected to the motor terminals so that the IR drops of the brushes at rated current are included. However, if some non-linear representation of IR drop for brushes is used for computer studies, the brushes should be short-circuited during the measurements for R_a . For the more accurate representation, the IR drop of the brushes as a function of armature current I_a should be measured and inserted in the applicable block diagram and computer representation.

14-2.16 Armature Inductance

Precise measurements of the armature inductance L_a are often unnecessary because this parameter does not affect appreciably the principal time constant of the motor. Probably the most convenient method of measuring L_a , and one of sufficient accuracy, requires an a-c voltmeter and an a-c ammeter. A 60-cps voltage is first applied to the armature which is blocked to prevent rotation; the field is excited at rated d-c field current. Measured

values of the voltage E across the armature and the armature current I are substituted with the known value of armature resistance R_a in the equation

$$X = \sqrt{\left(\frac{E}{I}\right)^2 - R_a^2} \quad (14-52)$$

where

X = reactance of armature inductance, in ohms

Once the reactance is known, the armature inductance can be calculated from the equation

$$L_a = \frac{X}{377} \quad (14-53)$$

An alternate method of measuring the armature inductance is to observe the transient growth of current when the armature is excited by a step voltage.⁽⁸⁾ The armature should be blocked to prevent rotation and the field should be excited at rated current. The step change in voltage is applied by closing a switch connected in series with the armature, a series resistance R_s , and a d-c source. The armature inductance can be calculated from the time-constant relationship of the voltage rise across the series resistance as observed by the aid of a cathode-ray oscilloscope; that is,

$$L_a = (T^*) (R_a + R_s) \quad (14-54)$$

where

T^* = observed time constant of armature and series resistance R_s , in sec

14-2.17 Field Resistance

In the absence of manufacturer's data, the field resistance R_f can generally be determined by measurement with an ordinary ohmmeter. If the resistance is too low to be read accurately on an ohmmeter, a Wheatstone or Kelvin bridge can be used.

14-2.18 Field Inductance

The field inductance L_f may be determined by either of the methods described to measure the armature inductance. Rated d-c armature current should be used to obtain more accurate measurements of L_f . If the inductance is determined by finding its reactance,

an a-c current whose rms value is about 1/2 the rated d-c current should be used. If the inductance is determined by the transient method, a step voltage that will result in a steady-state field current of 1/2 to 3/4 the rated value should be used.

14-2.19 Motor Constant

The motor constant K may be determined from a measurement of the stall torque at a known armature current and a known field current. By solving Eq. (14-2) for K , the motor constant is

$$K = \frac{T}{I_f I_a} \quad (14-55)$$

The stall torque should be measured at various angular positions of the armature, and an average of the maximum and minimum values should be used. If the motor is used with a constant field input, the value of normal field current should be used during the test.

A second method of determining the motor constant is to measure the armature current and speed for a given terminal voltage V_t . Knowing the armature resistance, the motor constant is then calculated from Eq. (14-1) by solving for K to obtain

$$K = \frac{V_t - I_a R_a}{I_a \omega} \quad (14-56)$$

14-2.20 Armature Inertia

If the armature can be easily removed from the motor, the armature inertia J can be measured by use of a piano-wire torsion pendulum. A small diameter wire, but one strong enough to support the armature, should be used. The longer the wire used, the lower the period of the pendulum, and the more accurate the measurements obtained. The pendulum should be calibrated first by measuring its period with a body of known (or calculable) inertia and simple geometry. Then, only the armature should be used; from the resulting period, the inertia is calculated by

$$J = J_c \left(\frac{T_f}{T_c} \right)^2 \quad (14-57)$$

where

J = unknown inertia of the armature, in kilogram-meters²

J_c = known inertia of the body used for calibration, in kilogram-meters²

T_f = period when armature is used, in sec

T_c = period when body of known inertia is used, in sec

The armature inertia can be determined by a retardation test without removing the armature from the motor. This test consists of measuring: (1) the time T_1 for the armature to coast from full speed to half speed when the armature power is removed; and (2) the time T_2 , obtained in the same manner, but when an additional body of known inertia is coupled to the armature. Let the ratio T_1/T_2 be T_R . Then

$$T_R = \frac{T_1}{T_2} = \frac{J}{J + J_a} \quad (14-58)$$

where

J_a = known inertia of body added to armature, in kilogram-meters²

Solving Eq. (14-58) for armature inertia J

$$J = J_a \left(\frac{T_R}{1 - T_R} \right) \quad (14-59)$$

If a tachometer is used to measure the motor speed during the retardation test, the tachometer inertia should be subtracted from the inertia found by Eq. (14-59) to obtain the true armature inertia. A satisfactory method of measuring motor speed without using the tachometer is with the d-c speed bridge.⁽¹²⁾

The armature inertia can be calculated from the following equation based on the inertia of a cylinder, if the armature dimensions and material are known:

$$J_{cyl} = 1/2 M r^2 \quad (14-60)$$

where

J_{cyl} = inertia of the cylinder, in in.-lb-sec²

$M = \frac{W}{g}$ = mass of the cylinder, in lb-sec²/in.

$$W = \pi \rho l r^2$$

ρ = weight density, in lb/in.³

l = cylinder length, in in.

r = cylinder radius, in in.

g = 386 in./sec²

14-2.21 Viscous Damping

The viscous damping B should be determined when the motor is operating without a load and at rated field input. The armature current should be measured when only enough voltage is applied to the armature to obtain rated motor speed. The total armature input power is then

$$V_t I_a = I_a^2 R_a + B \omega^2 \quad (14-61)$$

Solving for the viscous damping results in

$$B = \frac{(V_t I_a - I_a^2 R_a)}{\omega^2} \quad (14-62)$$

where

B = viscous damping, in newton-meter-sec

The measurement should be repeated at 1/2 rated speed and the two results averaged to obtain a more accurate value for B .

14-2.22 Typical Parameter Values

The typical parameter values of eight d-c motors are listed in Table 14-1. This table can be used as a guide to indicate the approximate values that can be expected. The "stall torque/inertia" quantity is a Figure of Merit indicating the acceleration that can be expected from the motor when unloaded. The motor time constant $R_a J / K^2$ characterizes the motor when the viscous damping B , armature inductance L_a , and load inertia are neglected. This simplification is often valid and useful in block diagram representations.

14-2.23 D-C TORQUE MOTORS

D-c torque motors are special motors whose output is a rotary displacement that is proportional to the control-winding input. To achieve linear operation, the output displacement is usually limited to an over-all rotation

of 10 to 20 degrees. As a result, torque motors are best adapted for the control of hydraulic control devices. In many applications, the rotary-displacement output is converted by a simple mechanical linkage into a translational displacement.

14-2.24 Description of Typical Torque Motors

The construction of a typical torque motor used to drive a hydraulic control valve is shown in Fig. 14-14A. The magnetic circuit for the polarizing flux ϕ_1 consists of permanent magnets $N-S$, magnetic yokes $Y-Y$, pole pieces $P-P$, rotor (armature) R , and the air gaps between the rotor and the pole pieces. Two control coils surround the rotor and are connected in push-pull (Fig. 14-14B) so that the magnetomotive forces of the control components i of the coil currents add to produce the control fluxes ϕ_c , but the magnetomotive forces of the quiescent current I_o cancel. When the coil currents are balanced (control components i are zero), the rotor is held in a neutral position by a spring. However, if the rotor is disturbed slightly or if the coil currents are unbalanced, the fluxes at one pair of diagonally opposite gaps increase, and the fluxes at the other pair of gaps decrease. The resultant force turns the rotor until equilibrium is reached against the spring or against stops. Without a spring, the rotor is unstable in the neutral position and, once disturbed, the rotor tends to clap immediately against the poles. Drive rods for hydraulic control valves are attached near the ends of the rotor at radius r . By using flexible drive rods, the rotational forces produced by the torque motor are converted into translational forces at the hydraulic control valves.

14-2.25 Static Characteristics

The torque developed by a typical torque motor (Fig. 14-14) is expressed by the equation

$$T = K_1 i + K_2 \theta_c \quad (14-63)$$

where

T = torque produced by motor, in lb-in.

i = control component of coil current, in amperes

TABLE 14-1 TYPICAL VALUES OF D-C MOTOR PARAMETERS

| Motor No. | Type | Nominal Size (hp) | Rated Voltage (volts) | Motor Constant (newton-meters/ampere) | Armature Resistance (ohms) | Inertia (kilogram-meters ²) | Stall Torque (newton-meters) | Stall Torque/Inertia (sec ⁻²) | Motor Time Constant $R_a J / K^2$ |
|-----------|------------------|-------------------|-----------------------|---------------------------------------|----------------------------|---|------------------------------|---|-----------------------------------|
| 1 | shunt | 0.008 | 115 | 0.23 | 280 | 1.1×10^{-5} | 0.095 | 8600 | 0.061 |
| 2 | permanent magnet | 0.003 | 6 | 0.011 | 2.15 | 0.24×10^{-5} | 0.025 | 10,200 | 0.043 |
| 3 | permanent magnet | 0.009 | 6 | 0.0091 | 0.82 | 0.37×10^{-5} | 0.062 | 17,000 | 0.036 |
| 4 | permanent magnet | 0.022 | 6 | 0.0093 | 0.38 | 0.53×10^{-5} | 0.141 | 27,000 | 0.023 |
| 5 | permanent magnet | 0.054 | 26 | 0.015 | 2.64 | 0.37×10^{-5} | 0.135 | 37,000 | 0.043 |
| 6 | permanent magnet | 0.080 | 26 | 0.0143 | 1.85 | 0.23×10^{-5} | 0.181 | 34,000 | 0.047 |
| 7 | shunt | 0.6 | 80* 100** | 0.198 | 5.82 | 2.9×10^{-4} | 2.72 | 9300 | 0.043 |
| 8 | shunt | 2.5 | 65* 110** | 0.329 | 1.57 | 1×10^{-3} | 13.6 | 13,200 | 0.016 |

*Armature voltage

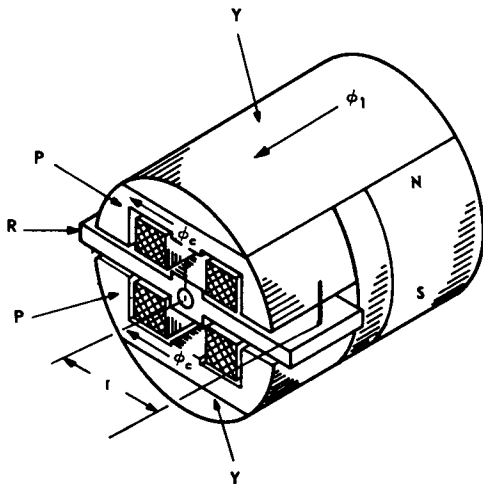
**Field voltage

θ_o = angular displacement of rotor, in radians

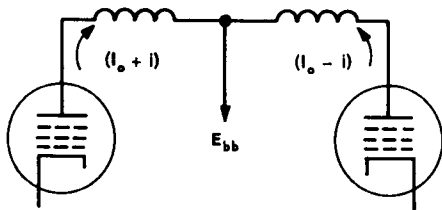
K_1 = constant (typical value is 200 lb-in./ampere)

K_2 = constant (typical value is 300 lb-in./radian)

The torque produced by the motor is opposed by a spring with constant K_s (typical value is 560 lb-in./radian). The static characteristic of the torque motor is shown in Fig. 14-15A, with the angular rotor deflection plotted as a function of the control current.



A. CROSS SECTION OF MAGNETIC CIRCUIT AND COILS



B. COIL CONNECTIONS

Fig. 14-14 D-c torque motor, permanent-magnet type.

Adapted from 'A Permanent-Magnet-Type Electric Actuator for Servovalves', by R. H. Frazier and R. D. Atchley, Dynamic Analysis and Control Laboratory, Report No. 66, June 1, 1952, Massachusetts Institute of Technology.

14-2.26 Dynamic Characteristics

The dynamic behavior of a typical torque motor (Fig. 14-14) is described by the equations

$$K_1 i = J_o s^2 \theta_o + B_o s \theta_o + (K_s - K_2) \theta_o \quad (14-64)$$

$$E_c = \mu e_1 = R_1 i + L_o s i + 0.1135 K_1 s \theta_o \quad (14-65)$$

where

J_o = inertia of load, in lb-in.-sec²

B_o = viscous damping of load, in lb-in.-sec

E_c = control-winding input voltage, in volts

μ = voltage gain of one tube in output stage of push-pull amplifier

e_1 = total voltage applied from grid-to-grid of amplifier output stage, in volts

$R_1 = 2r_p + R$ = total control-circuit resistance, in ohms

r_p = plate resistance of one tube in amplifier output stage, in ohms

R = total resistance of control winding, in ohms

L_o = self-inductance of control winding (two coils in series and rotor in neutral position), in henries

K_1 = torque per unit current, in lb-in./ampere

The dynamic behavior of the torque motor is summarized in terms of the frequency response as shown in Fig. 14-15B. An indication of the frequency response of other torque motors used to drive hydraulic control valves may be obtained from the information tabulated for hydraulic-amplifier dynamic response in Par. 13-6.

14-2.27 Modified Servomotors

Most servomotors can be modified to serve as a torque motor by adding a gear train to provide the small motions usually required of the torque motor. The motor and gear train can be converted into a position source by opposing the motor torque with that of a

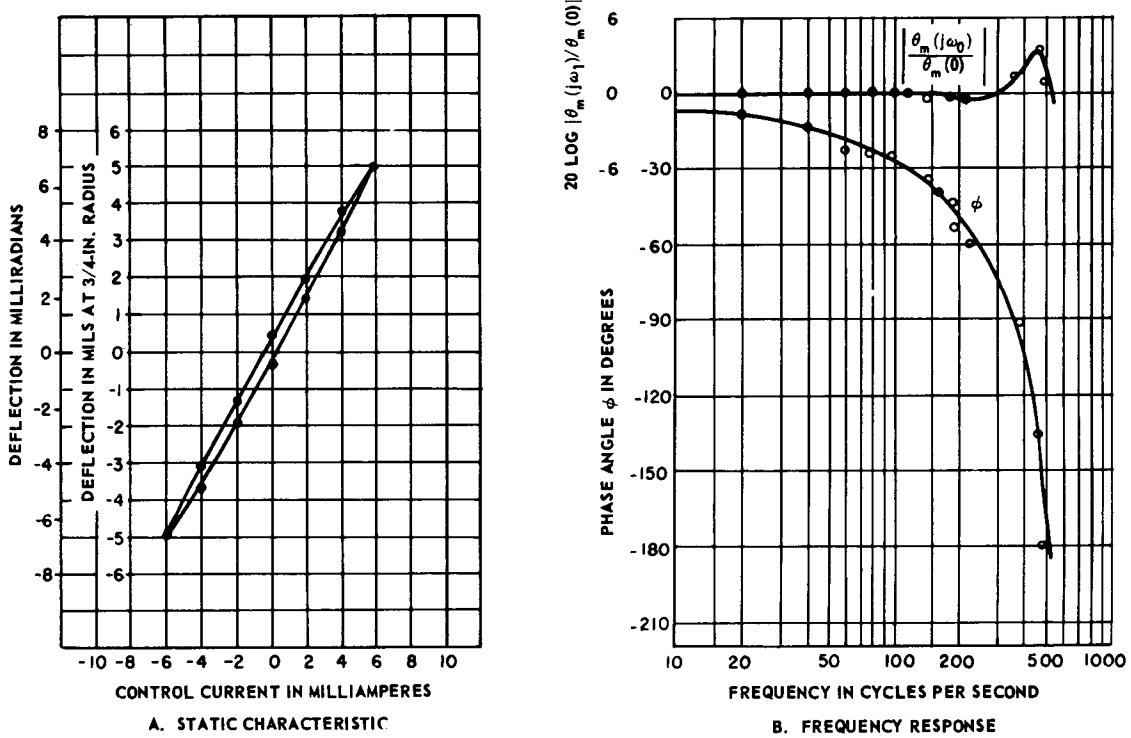


Fig. 14-15 Characteristics of d-c torque motor.

From 'A Permanent-Magnet-Type Electric Actuator for Servovalves', by R. H. Frazier and R. D. Atchley, Dynamic Analysis and Control Laboratory, Report No. 66, June 1, 1952, Massachusetts Institute of Technology.

spring. In this case, the steady-state behavior is

$$\theta_o = \frac{R_g T_m}{K_s} \quad (14-66)$$

$$T_m = K_m E_c \quad (14-67)$$

where

θ_o = angular displacement of motor-spring combination, in radians

R_g = gear reduction ratio of gear train

T_m = torque produced by motor, in lb-in.

K_s = spring constant of opposing spring, in lb-in./radian

K_m = motor torque per volt constant, in lb-in./volt

E_c = control-winding input voltage, in volts

Equations (14-66) and (14-67) apply to a d-c motor with E_c applied to the armature and with a constant field input; to a d-c motor with E_c applied to the field and with a constant armature input; or to a 2-phase a-c servomotor.

14-2.28 CONVERSION FACTORS AND UNITS

The conversion factors and units commonly required when making calculations and measurements to determine the performance of d-c motors are listed in Table 14-2.

POWER ELEMENTS AND SYSTEM DESIGN

TABLE 14-2 CONVERSION FACTORS AND UNITS

| Quantity | To Convert English Units | To Metric (Mks) Units | Multiply By |
|---------------------------|-----------------------------|------------------------------|----------------|
| mass (M) | lb-sec ² /in. | kilograms | 175 |
| inertia (J) | lb-in.-sec ² | kilogram-meters ² | 0.113 |
| force (F) | lb | newtons | 4.45 |
| torque (T) | lb-in. | newton-meters | 0.113 |
| viscous damping (B) | lb-in.-sec | newton-meter-sec | 0.113 |
| spring constant (K_s) | lb-in. | newton-meters | 0.113 |

APPENDIX FOR PARAGRAPH 14-2.

Solve Eq. (14-1) and (14-2) for torque. There results

$$T = \frac{KI_f V_t}{R_a} - \frac{KI_f V_t}{R_a} \times \frac{KI_f \omega}{V_t} \quad (1)$$

With zero speed and maximum current, I_{fM} , the stall torque T_M is seen to be $KI_{fM}V_t/R_a$. Divide the expression for T from Eq. (1) by T_M on the left and the equivalent expression on the right. There results.

$$\frac{T}{T_M} = \frac{I_f}{I_{fM}} - \frac{I_f}{I_{fM}} \times \frac{KI_f \omega}{V_t} \quad (2)$$

Solve Eq. (1) for the no-load speed ω_n , at current I_{fM} , which is given by:

$$\frac{KI_{fM}\omega_n}{V_t} = 1 \quad (3)$$

Rewrite Eq. (2) as follows:

$$\frac{T}{T_M} = \frac{I_f}{I_{fM}} \left(1 - \frac{KI_f}{I_{fM}} \times \frac{\omega_n I_{fM}}{V_t} \times \frac{\omega}{\omega_n} \right) \quad (4)$$

Substitute Eq. (3) into Eq. (4) to arrive at the following [which is Eq. (14-14)]

$$\frac{T}{T_M} = \frac{I_f}{I_{fM}} \left(1 - \frac{I_f}{I_{fM}} \times \frac{\omega}{\omega_n} \right)$$

POWER ELEMENTS USED IN CONTROLLERS

14-3 ALTERNATING-CURRENT MOTORS

14-3.1 TYPES OF A-C MOTORS USED IN SERVOMECHANISMS

The following three types of a-c motors are commonly used as servomotors (output members of servomechanisms) :

- (a) 2-phase motors (commonly used in continuous linear servomechanisms)
- (b) Single-phase motors (commonly used in relay servomechanisms)
- (c) Torque motors with restricted rotation (commonly used to control the input to hydraulic amplifiers)

Amplification for driving servomotors may be furnished by magnetic, transistor, or vacuum-tube amplifiers (see Ch. 13). Typical schematic diagrams showing the use of servomotors and amplifiers are included in this paragraph.

14-3.2 Amplifier Control of Servomotors

Figures 14-16 through 14-19 show the

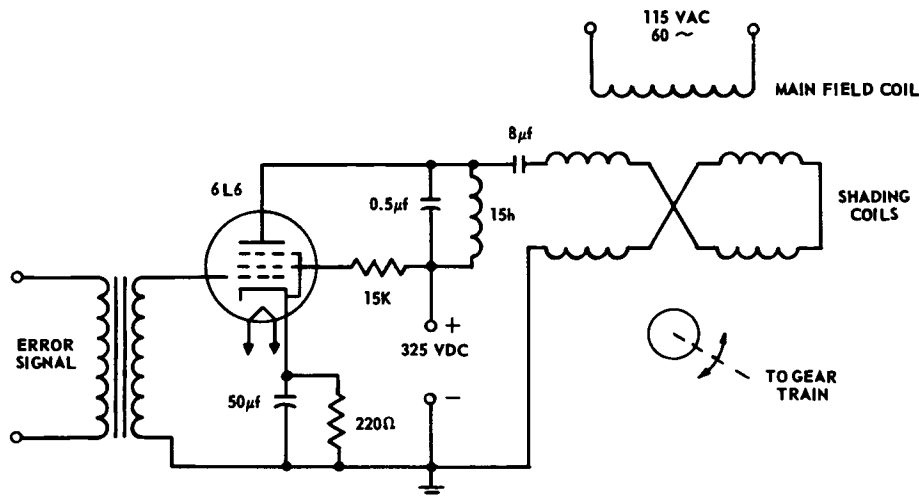
schematic diagrams of a few servomotor control circuits using amplifiers:

(a) Figure 14-16 shows a shaded-pole motor in a 60-cycle servo application. Wound shading coils are used, and the main winding is continuously excited.

(b) Figure 14-17 shows a 2-phase 60-cycle servomotor controlled by a vacuum-tube amplifier. A feedback loop is used to reduce the amplifier internal impedance presented to the motor winding, thus increasing the inherent motor damping. A high amplifier internal impedance would have the opposite effect; in some cases, excessive impedance can reduce the damping to zero or even make it negative. (For 400-cycle operation, the size of amplifier coupling capacitors must be changed.)

(c) Figure 14-18 shows a 2-phase servomotor controlled by power transistors.

(d) Figure 14-19 shows several magnetic amplifier circuits proposed by Geyger.⁽¹³⁾ The



NOTE: ADDITIONAL AMPLIFICATION AND COMPENSATION MAY BE REQUIRED.

Fig. 14-16 Amplifier control circuit for shaded-pole motor.

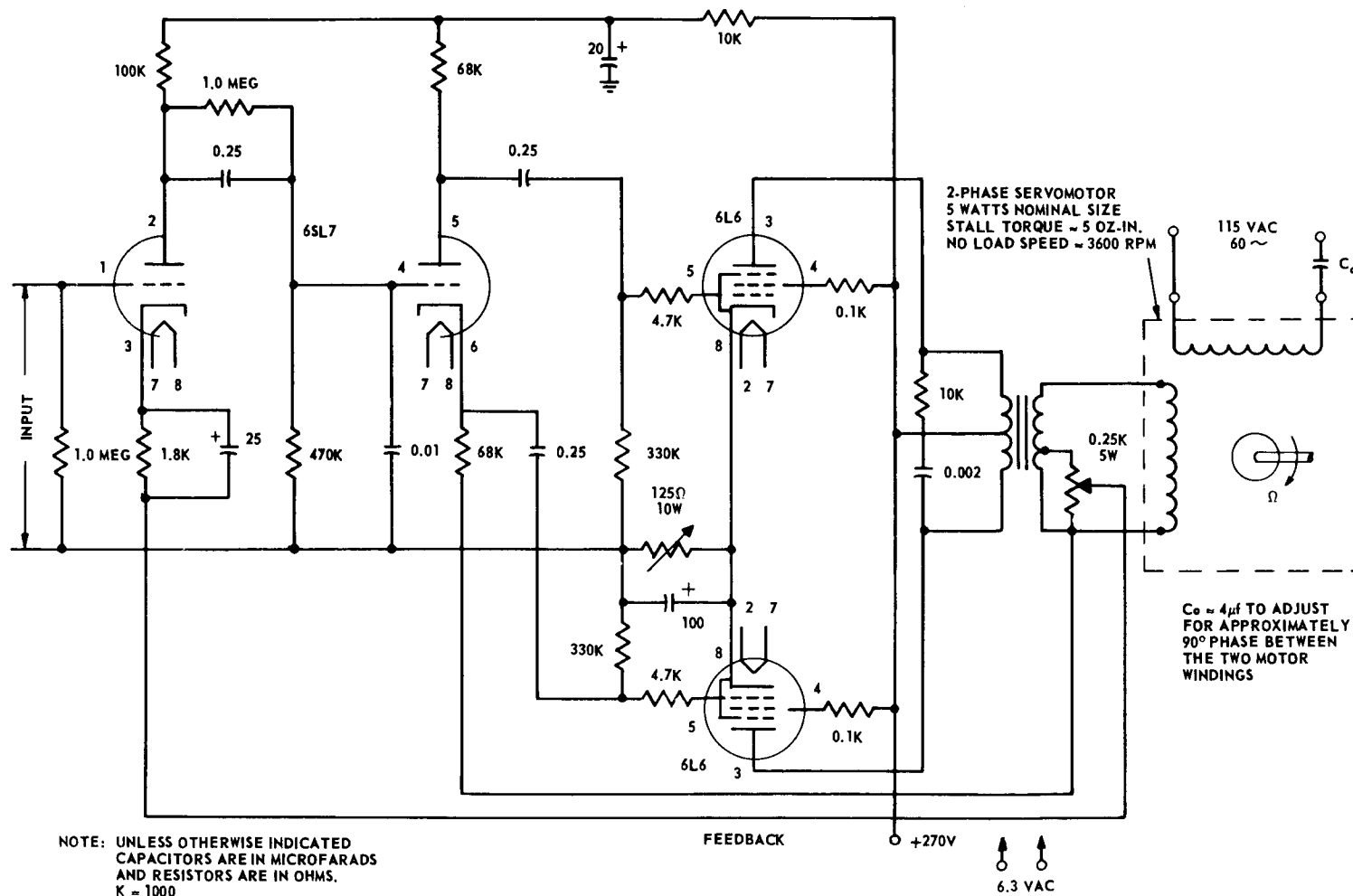


Fig. 14-17 Amplifier control circuit for 2-phase servomotor.

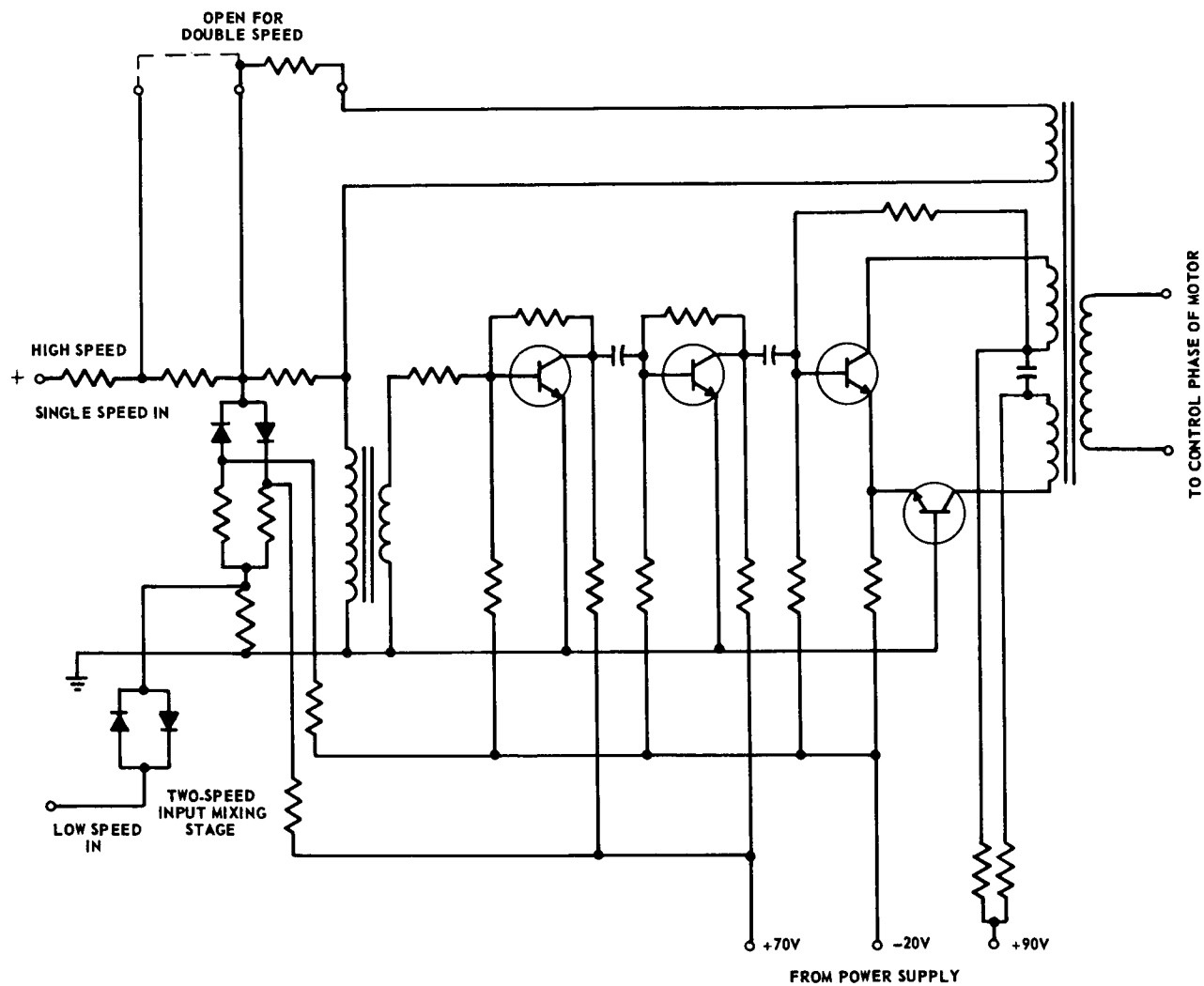


Fig. 14-18 Schematic of a transistor amplifier used with a 2-phase servomotor.

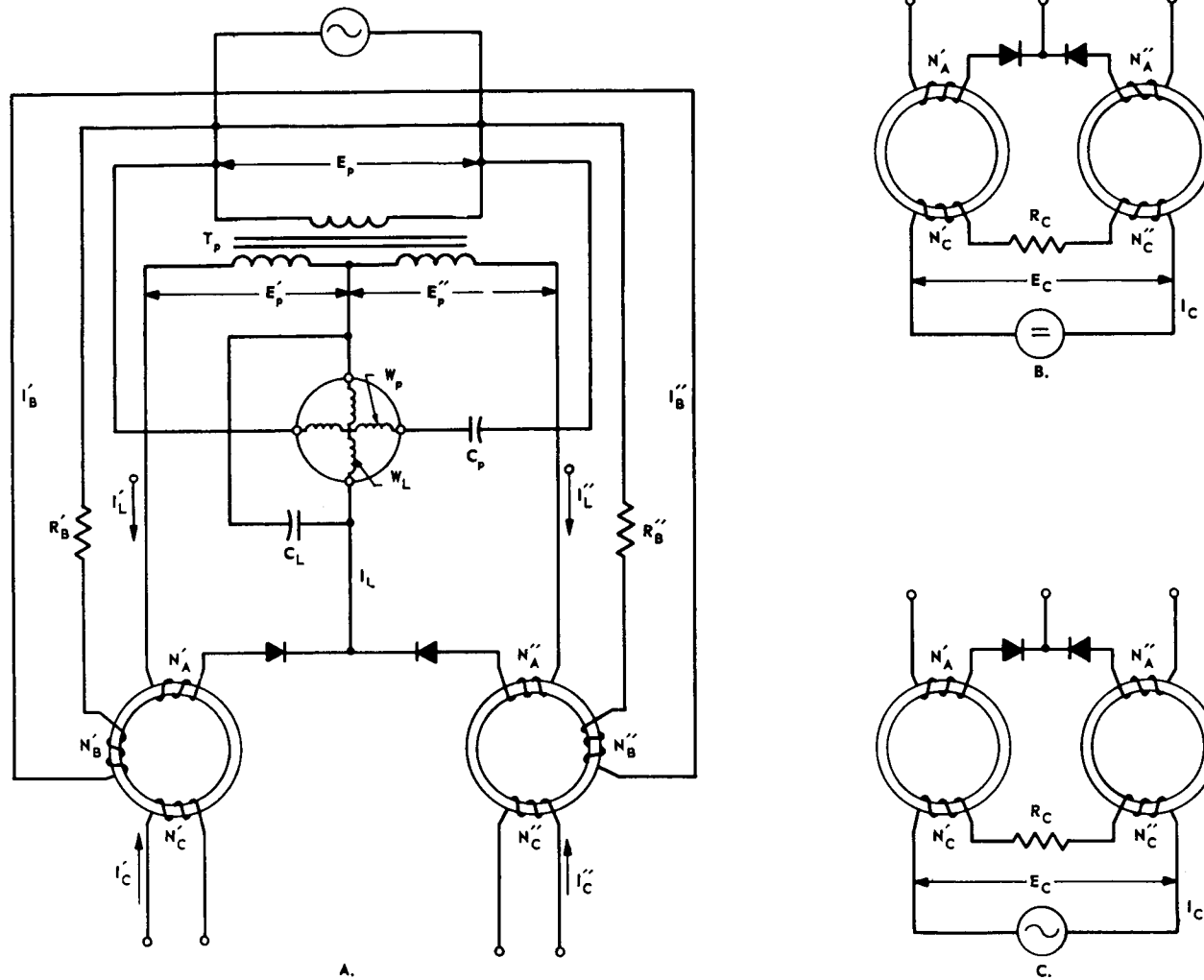


Fig. 14-19 Full-wave output-stage circuit of a magnetic servo amplifier having inherent dynamic-braking properties: (A) with two separate control winding elements $N'_c N''_c$, (B) with series-connected d-c control winding elements $N'_c N''_c$, (C) with series-connected a-c control winding elements $N'_c N''_c$.

CHAPTER 15

MECHANICAL AUXILIARIES USED IN CONTROLLERS***15-1 GEAR TRAINS****15-1.1 PURPOSE**

Gear trains are most frequently used in controllers to transmit motion from the servomotor to the load with an appropriate change in speed. The transmitted power can range from a small fraction of a horsepower for instrument-type servomechanisms to several horsepower for large gun or launcher drives.

Gear trains are also used to transmit input, feedback, and error-signal information, when an advantage will be gained by this use of gear trains, as in a completely mechanical-hydraulic servomechanism. In most cases, such gear trains transmit very small amounts of power.

The discussion of gear trains will be confined to gears with involute tooth form, because this tooth form is used almost exclusively for engineering purposes. Other tooth forms which result in rolling action between meshing teeth have been made, but these special tooth forms are confined to very special applications, and cutters for their manufacture are not normally available.

15-1.2 DEFINITIONS

The most important parts of a gear are defined in the following subparagraphs (Fig. 15-1) :

*By J. O. Silvey

(a) The *pitch line or pitch circle* of a gear is that circle which has the same tangential velocity as the pitch circle of a second gear, when the two gears are properly meshed and rotated together. The pitch circle is located slightly closer to the tops of the gear teeth than it is to the bottoms of the spaces between teeth.

(b) The *pitch diameter* of a gear is the diameter of the pitch circle. In bevel gears, it is the largest diameter of the pitch cone.

(c) The *circular pitch* is the distance (normally expressed in inches) on the circumference of the pitch circle between corresponding points of adjacent teeth. Circular pitch is equal to π times the pitch diameter divided by the number of teeth on the gear.

(d) The *diametric pitch* is equal to the number of teeth on the gear divided by the pitch diameter.

(e) The *addendum* is the radial distance between the pitch circle and the tops of the teeth.

(f) The *dedendum* is the radial distance between the pitch circle and the bottoms of the spaces between the gear teeth. The dedendum of a gear is always slightly larger than the addendum to provide clearance for the cutter of generated gear teeth, and for the tops of the teeth of a mating gear.

(g) The *face width* is the length of the teeth along the axis of the gear.

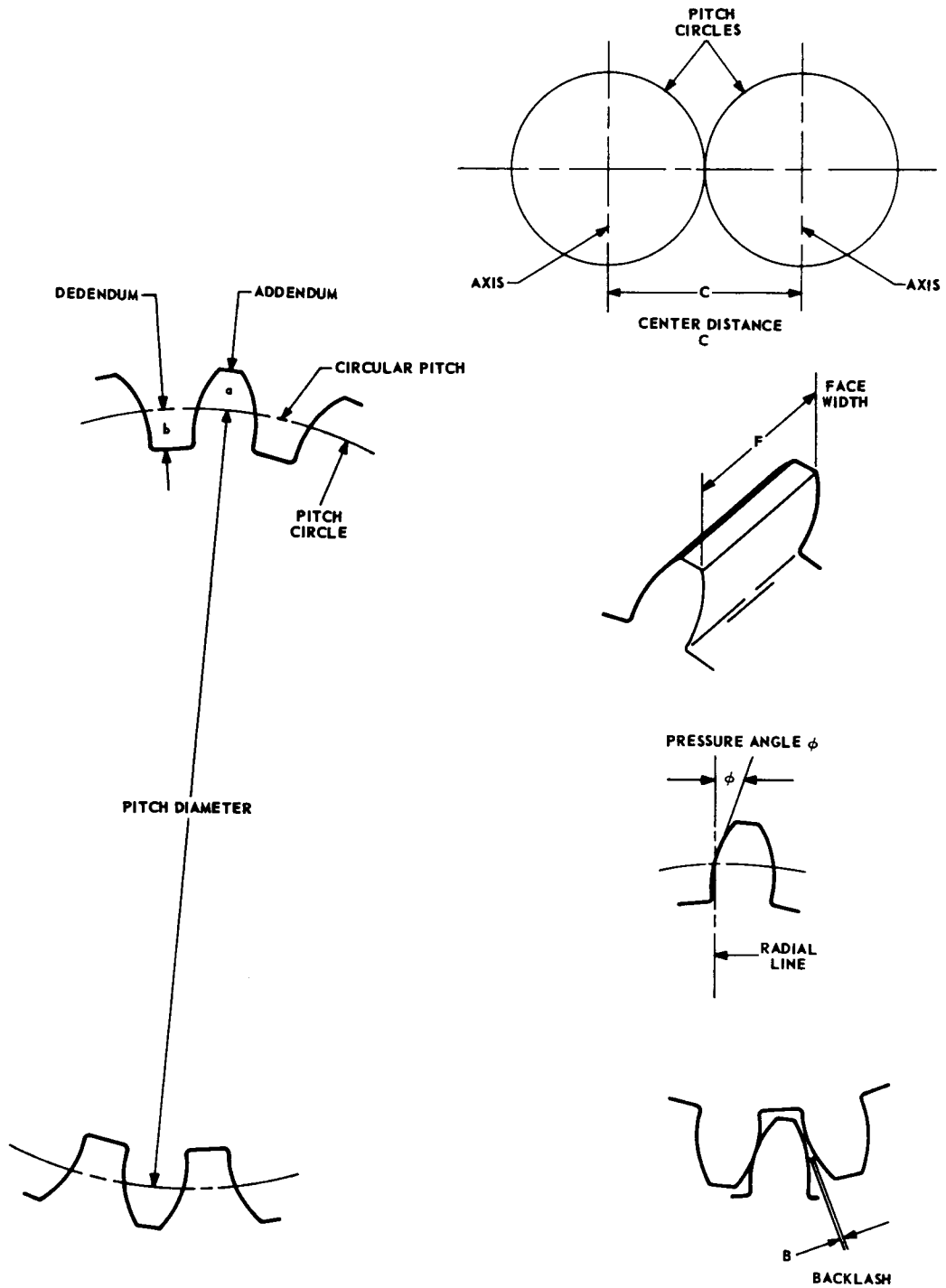


Fig. 15-1 Most important parts of a gear.

(h) The *pressure angle* is the angle between a tangent to the tooth profile at the pitch circle and a radius to this pitch point. Pressure angle is also the angle of inclination of the sides of the teeth in the basic rack.

(i) The *center distance* is the shortest distance between gear axes of a pair of mating gears. For gear pairs with parallel shafts, the center distance is the distance between axes.

(j) The *backlash* is the play between mating teeth or the shortest distance between the nondriving surfaces of adjacent teeth.

(k) The *gear ratio* of a pair of mating gears is the ratio of the number of teeth on the gear divided by the number of teeth on the pinion.

15-1.3 GEAR TYPES

The following five types of gears are in common use (Fig. 15-2) :

(a) *Spur gears* (Fig. 15-2A) are used more in control work than any other type gear. They are the least expensive to manufacture, have high mechanical efficiency, and, except where the gear ratio is high, can be operated to either increase or decrease speed. Spur gear teeth are parallel to the gear axis, producing negligible thrust along their shafts when driven. Four spur-gear systems will satisfy most requirements. They are classified by pressure angle and tooth form, and have been standardized by the American Standards Association.⁽¹⁾ These four systems are as follows :

(1) *14-1/2-deg composite*. Gears of this system are machined by milling one tooth space at a time, a series of eight cutters being required to cover the range between a 12-tooth gear and rack. Gears of the other three systems are machined by a hob or shaper.

(2) *14-1/2-deg full-depth tooth*. Gears of this system are usually produced by hobbing and have the same strength as gears of the composite system. However, full-depth

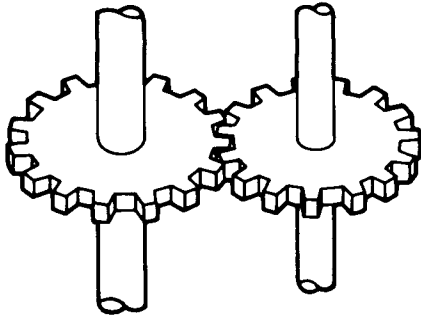
pinions of standard proportions but with less than 32 teeth will have the teeth undercut, and the undercut may become excessive for smooth operation of gears with less than 22 teeth.

(3) *20-deg full-depth tooth*. This system produces stronger gear teeth than the 14-1/2 degree system. Teeth are produced by hobbing or shaping and are not undercut in gears with more than 18 teeth. Undercut may be excessive in gears with less than 14 teeth.

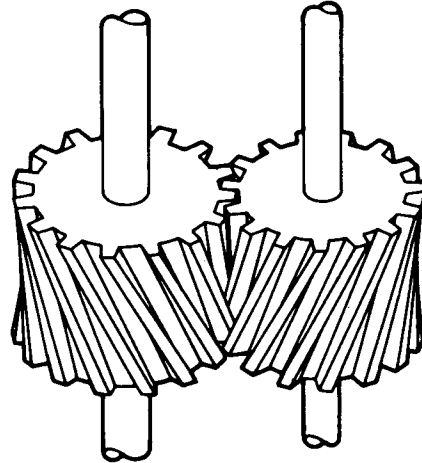
(4) *20-deg stub tooth*. This system produces the strongest teeth, which are hobbled or shaped. It also produces pinions with the least number of teeth without undercut. By increasing the diameter of the pinion above that specified by standard gear proportions, the undercut of small pinions can be reduced. The center distance of the gear pair must be increased to accommodate the larger-diameter pinion.

(b) *Helical gears*. The teeth of helical gears (Fig. 15-2B) are cut in the form of a helix on the pitch cylinder, the helix angle variable between widely separated limits. Some end thrust is developed when helical gears are driven. Helical gears operate quieter, and produce smoother action than spur gears of equal precision. However, helical gears are more expensive to produce. When helical gears are mounted with their shafts at right angles to each other, they are called spiral gears, or crossed helical gears. Two side-by-side helical gears, with equal pitch diameters and equal but opposite helical angles, comprise a herringbone gear. Such a gear can be cut in a single piece of stock, or it can comprise two separate gears mounted together (side-by-side as a unit) on the same shaft. Herringbone gears transmit no thrust to their bearings because the thrusts of the two helical gears are equal but act in opposite directions.

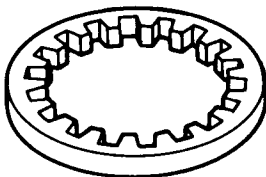
(c) *Internal gears*. The primary uses for internal gears (Fig. 15-2C) are as clutches, splines, and as components of planetary-gear systems.



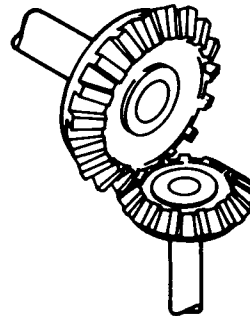
A. SPUR



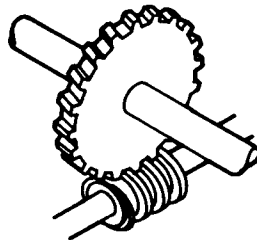
B. HELICAL



C. INTERNAL



D. BEVEL



E. WORM AND WORM GEAR

Fig. 15-2 Types of gears.

(d) *Bevel gears.* These gears (Fig. 15-2D) are mounted with their shafts at an angle to each other. When the gear ratio is unity, and the shafts are at an angle of 90° , bevel gears are called mitre gears. The teeth of bevel gears may be straight or helical, either type always producing end thrust when driven. To change the gear ratio of bevel gears, both gears of a pair must be changed.

(e) *Worms and worm gears.* Worms and worm gears (Fig. 15-2E) are mounted with their shafts at right angles to each other. The worm may have a single thread or several threads. Worm-and-gear systems are often used to obtain a large speed reduction in a small space, but the mechanical efficiency is low because the action between teeth is of a sliding nature, instead of rolling. Normally, worm drives can only be driven by rotating the worm, not by rotating the gear, because of the low helix angle of the worm.

15-1.4 DESIGN FUNDAMENTALS⁽³⁾

The subsequent paragraphs cover in some detail the following design fundamentals: backlash, dynamic load, gear accuracy, beam strength of teeth, Y factor, margin of safety, and limit load for wear. Equations and tables are supplied to supplement the descriptive text. The presented method of computing permissible gear load is adequate for the great majority of gearing problems encountered in control systems. If extremely small size or high speed is required in a gear train, other methods, such as those employed by Almen,⁽²⁷⁾ McFarland,⁽²⁸⁾ and others should be used. In general, the design of such gears requires a great deal of experience and should be done by a gear expert.

15-1.5 Backlash

Backlash between meshed gears is the gap or clearance required between nondriving surfaces of adjacent teeth to prevent binding. High-speed gearing requires more backlash than low-speed gearing. If appreciable heat is generated by the gears, enough additional backlash must be provided to permit thermal expansion of the gear material. In any gear-train design, the backlash between

gear pairs must also be great enough to prevent binding due to eccentricities of the bearings, shafts, and gear-pitch circles. Backlash standards for general-purpose spur gearing have been established by the American Standards Association,^(2,3,4) and a method of specifying backlash in fine-pitch gearing has also been adopted.⁽⁵⁾

The required amount of backlash can be obtained by cutting the teeth thinner than the theoretical optimum dimensions. This is achieved by cutting the teeth deeper than normal. In most cases, the teeth of each gear of a mating pair are cut thinner by an amount equal to half the required backlash to retain as much strength as possible in each gear. However, where small pinions are used, all of the backlash should be obtained by cutting the teeth of the mating gear thinner by an amount equal to the total required backlash. The excess depth of cut X to provide the required backlash when both gears are cut deeper is

$$X = \frac{B}{4 \sin \phi} \quad (\text{for both gears}) \quad (15-1)$$

and, when only one gear is cut deeper

$$X = \frac{B}{2 \sin \phi} \quad (\text{one gear only}) \quad (15-2)$$

where

B = amount of backlash, in in.

X = excess depth of cut to provide backlash, in in.

ϕ = pressure angle of generating tool

15-1.6 Dynamic Load

The maximum momentary load set up by the operating or dynamic conditions is called the dynamic load and is given by

$$W_d = \frac{0.05V(FC + W)}{0.05V + \sqrt{FC + W}} + W \quad (15-3)$$

where

$$W = \frac{33,000 HP}{V} = \text{total applied load, in lb} \quad (15-4)$$

W_d = total dynamic load, in lb

POWER ELEMENTS AND SYSTEM DESIGN

V = pitch-line velocity, in fpm

HP = number of horsepower transmitted

C = deformation factor (see Table 15-1)

F = face width of gears, in in.

For planetary and differential gear trains, the velocity of actual tooth engagement must be used to determine the amount of dynamic load. For materials not listed in Table 15-1, the value of deformation factor C may be determined as follows:

$$C = \frac{0.107e}{\frac{1}{E_1} + \frac{1}{E_2}} \quad (\text{for } 14\text{-}1/2\text{-deg tooth form}) \quad (15-5)$$

$$C = \frac{0.111e}{\frac{1}{E_1} + \frac{1}{E_2}} \quad (\text{for } 20\text{-deg full depth form}) \quad (15-6)$$

$$C = \frac{0.115e}{\frac{1}{E_1} + \frac{1}{E_2}} \quad (\text{for } 20\text{-deg stub tooth form}) \quad (15-7)$$

where

e = error in action, in in.

E_1 and E_2 = modulus of elasticity of materials

To solve Eqs. (15-5) through (15-7) the error in action e of the gears must be known. Judging from test results, the error in action on well-cut commercial gears ranges from about 0.002 inch on gears of 6 dp (diametric pitch) and finer, to about 0.004 inch on gears of 2 dp. When the gears are cut with great care, these errors are reduced. On ground gears made with extreme care, as well as on cut gears where every effort is made to achieve the utmost accuracy, the errors are

TABLE 15-1 VALUES OF DEFORMATION FACTOR C

| Material | Tooth Form (deg) | Error in Action (in.) | | | | | |
|-------------------------|---------------------|-----------------------|-------|-------|-------|-------|-------|
| | | 0.0005 | 0.001 | 0.002 | 0.003 | 0.004 | 0.005 |
| Cast iron and cast iron | 14-1/2 | 400 | 800 | 1600 | 2400 | 3200 | 4000 |
| Cast iron and steel | 14-1/2 | 550 | 1100 | 2200 | 3300 | 4400 | 5500 |
| Steel and steel | 14-1/2 | 800 | 1600 | 3200 | 4800 | 6400 | 8000 |
| Cast iron and cast iron | 20 (full depth) | 415 | 830 | 1660 | 2490 | 3320 | 4150 |
| Cast iron and steel | 20 (full depth) | 570 | 1140 | 2280 | 3420 | 4560 | 5700 |
| Steel and steel | 20 (full depth) | 830 | 1660 | 3320 | 4980 | 6640 | 8300 |
| Cast iron and cast iron | 20 (stub) | 430 | 860 | 1720 | 2580 | 3440 | 4300 |
| Cast iron and steel | 20 (stub) | 590 | 1180 | 2360 | 3540 | 4720 | 5900 |
| Steel and steel | 20 (stub) | 860 | 1720 | 3440 | 5160 | 6880 | 8600 |

Adapted from *Manual of Gear Design*, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

MECHANICAL AUXILIARIES USED IN CONTROLLERS

reduced to even smaller amounts. The foregoing conditions of error in action are listed in Table 15-2 and are designated as Classes 1, 2, and 3.

15-1.7 Gear Accuracy

The noise of gear operation is usually a very good indication of gear accuracy. Table 15-3 gives some idea of the order of accuracy

required at different pitch-line velocities, and should serve as a guide for selection of the proper class of gear to meet specified speed conditions. Where extreme quietness of gear operation is required, a higher order of accuracy than shown in the table will be required. However, the values in the table should keep operating noise and the intensity of the dynamic load within reasonable limits.

TABLE 15-2 MAXIMUM ERROR IN ACTION BETWEEN GEARS AS A FUNCTION OF CLASS

| Diametric Pitch | Class 1 | Class 2 | Class 3 |
|-----------------|---------|---------|---------|
| 1 | 0.0048 | 0.0024 | 0.0012 |
| 2 | 0.0040 | 0.0020 | 0.0010 |
| 3 | 0.0032 | 0.0016 | 0.0008 |
| 4 | 0.0026 | 0.0013 | 0.0007 |
| 5 | 0.0022 | 0.0011 | 0.0006 |
| 6 and finer | 0.0020 | 0.0010 | 0.0005 |

Adapted from *Manual of Gear Design*, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

TABLE 15-3 MAXIMUM ERROR IN ACTION BETWEEN GEARS AS A FUNCTION OF PITCH LINE VELOCITY

| V (ft/min) | Error (in.) | V (ft/min) | Error (in.) | V (ft/min) | Error (in.) | V (ft/min) | Error (in.) |
|---------------|----------------|---------------|----------------|---------------|----------------|---------------------|----------------|
| 250 | 0.0037 | 1500 | 0.0019 | 2750 | 0.0010 | 4500 | 0.0006 |
| 500 | 0.0032 | 1750 | 0.0017 | 3000 | 0.0009 | 5000 and over | 0.0005 |
| 750 | 0.0028 | 2000 | 0.0015 | 3250 | 0.0008 | | |
| 1000 | 0.0024 | 2250 | 0.0013 | 3500 | 0.0007 | | |
| 1250 | 0.0021 | 2500 | 0.0012 | 4000 | 0.0006 | | |

Adapted from *Manual of Gear Design*, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

15-1.8 Beam Strength of Teeth

To calculate the beam strength of the gear teeth, the Lewis equation can be used to determine the safe static strength.

$$W_b = s_t p F y \quad (15-8)$$

where

W_b = safe static beam load on teeth, in lb

p = circular pitch, in in.

F = face width, in in.

s_t = safe static bending stress for materials, in psi

y = tooth form factor

Values of y for various tooth forms are given in Table 15-4. The flexural endurance limit of a gear material will give a satisfactory value for the safe static bending stress s_t . Fatigue tests on steel indicate that this endurance limit follows the Brinnell hardness number quite closely. Endurance limits are listed in Table 15-5. The static beam strength of the gear tooth should always be greater than the dynamic load; thus

$$W_b = 1.25 W_d \text{ for steady loads} \quad (15-9)$$

$$W_b = 1.35 W_d \text{ for pulsating loads} \quad (15-10)$$

$$W_b = 1.50 W_d \text{ for shock loads} \quad (15-11)$$

where

W_b = safe static beam load on teeth, in lb

W_d = total dynamic load, in lb

These values should be modified when found desirable by experience.

15-1.9 Limit Load for Wear

The limit load for wear depends upon the surface endurance limits of the materials, the radii of the curvature of mating profiles, and upon the relative hardness of the mating surfaces. A harder material in the pinion will cold-work the surface of the softer and more

malleable material in the mating gear, and will thus materially increase the surface endurance limit of the gear. The value of the limiting static load for wear should be equal to or greater than the value of the dynamic load, and is given by

$$W_w = D F K Q \quad (15-12)$$

where

W_w = limiting static load for wear, in lb

D = pitch diameter of pinion, in in.

F = face width, in in.

K = load-stress factor

Q = ratio factor

For spur gears

$$Q = \frac{2N_g}{N_p + N_g} \quad (15-13)$$

For internal gears

$$Q = \frac{2N_g}{N_g - N_p} \quad (15-14)$$

$$K = \frac{s_c' \sin \phi}{1.400} \left(\frac{1}{E_1} + \frac{1}{E_2} \right) \quad (15-15)$$

where

N_g = number of teeth in gear

N_p = number of teeth in pinion

ϕ = pressure angle of gears

s_c' = surface endurance limit of material, in psi

E_1 and E_2 = modulus of elasticity of materials

Values of the load-stress factor K are given in Tables 15-6 and 15-7 for various combinations of materials, taking into consideration the cold-working received in operation. The classification "cast iron" also includes the ordinary semi-steel. Some of the high-test and semi-steels, and other special alloys of cast iron, have a greater modulus of elasticity than cast iron. For these materials, the value of K and the value of the dynamic load W_d for such materials must be calculated directly

MECHANICAL AUXILIARIES USED IN CONTROLLERS

TABLE 15-4 VALUES OF TOOTH FORM FACTOR y FOR VARIOUS TOOTH FORMS

| Number of Teeth | 14-1/2° Composite and Involute Form | 20° Full Depth Involute System | 20° Stub Tooth System |
|-----------------|-------------------------------------|--------------------------------|-----------------------|
| 12 | 0.067 | 0.078 | 0.099 |
| 13 | 0.071 | 0.083 | 0.103 |
| 14 | 0.075 | 0.088 | 0.108 |
| 15 | 0.078 | 0.092 | 0.111 |
| 16 | 0.081 | 0.094 | 0.115 |
| 17 | 0.084 | 0.096 | 0.117 |
| 18 | 0.086 | 0.098 | 0.120 |
| 19 | 0.088 | 0.100 | 0.123 |
| 20 | 0.090 | 0.102 | 0.125 |
| 21 | 0.092 | 0.104 | 0.127 |
| 22 | 0.093 | 0.105 | 0.129 |
| 24 | 0.095 | 0.107 | 0.132 |
| 26 | 0.098 | 0.110 | 0.135 |
| 28 | 0.100 | 0.112 | 0.137 |
| 30 | 0.101 | 0.114 | 0.139 |
| 34 | 0.104 | 0.118 | 0.142 |
| 38 | 0.106 | 0.122 | 0.145 |
| 43 | 0.108 | 0.126 | 0.147 |
| 50 | 0.110 | 0.130 | 0.151 |
| 60 | 0.113 | 0.134 | 0.154 |
| 75 | 0.115 | 0.138 | 0.158 |
| 100 | 0.117 | 0.142 | 0.161 |
| 150 | 0.119 | 0.146 | 0.165 |
| 300 | 0.122 | 0.150 | 0.170 |
| Rack | 0.124 | 0.154 | 0.175 |

Adapted from *Manual of Gear Design*, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

TABLE 15-5 VALUES OF SAFE STATIC BENDING STRESS s_t (psi)

| Material | Brinnell Hardness Number | s_t |
|--------------|--------------------------|---------|
| Gray iron | 160 | 12,000 |
| Semi-steel | 190 | 18,000 |
| Phos. bronze | 100 | 24,000 |
| Steel | 150 | 36,000 |
| Steel | 200 | 50,000 |
| Steel | 240 | 60,000 |
| Steel | 280 | 70,000 |
| Steel | 320 | 80,000 |
| Steel | 360 | 90,000 |
| Steel | 400 | 100,000 |

Manual of Gear Design, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

from their own specific properties. The surface endurance limit of gear material is a measure of the load limit in psi that the material will tolerate before deformation occurs. The values of surface endurance limit s_e for steel appear to vary quite consistently with the Brinnell hardness number up to about 400 Brinnell hardness. Values of s_e for hardness numbers from 150 to 400 are listed in Table 15-6. When the Brinnell hardness number is much higher than 400, the steel does not appear to have a definite endurance limit. The values given in Table 15-7 are suggested for use with steels harder than 400 Brinnell hardness. Table 15-7 also gives values for the load-stress factor K for these harder steels. The load-carrying ability of a pair of metal spur gears, for example, may be limited by either the beam strength of the gear tooth, or by the surface endurance limit of the material. The lower of these two values should be used to establish the load-carrying ability of any given pair of gears.

Spur gears designed to conform to the preceding requirements can be expected to operate successfully with average loads. However, if heavy inertia loads are to be coupled rigidly to the gears, the dynamic load should be calculated in accordance with the method described by Buckingham.^(3,8)

Other types of gears such as internal, helical, bevel, and worm gears are used so infrequently in control work that the particulars of their design are not included here. If use of any of these types is contemplated, the following references should be consulted:

- (a) Internal gears — Buckingham⁽³⁾ and Kent⁽⁴⁾
- (b) Helical gears — Kent⁽⁴⁾ and Buckingham⁽⁶⁾
- (c) Bevel gears — Jones⁽⁷⁾
- (d) Worm gears — Kent,⁽⁴⁾ Buckingham,⁽⁶⁾ and Jones

MECHANICAL AUXILIARIES USED IN CONTROLLERS

TABLE 15-6 VALUES OF LOAD-STRESS FACTOR K FOR VARIOUS MATERIALS

| Material in Pinion | Brinnell Number | Material in Gear | Brinnell Number | s_c | K (14-1/2°) | K (20°) |
|-----------------------|--------------------|---------------------|--------------------|---------|------------------|--------------|
| Steel | 150 | Steel | 150 | 50,000 | 30 | 41 |
| Steel | 200 | Steel | 150 | 60,000 | 43 | 58 |
| Steel | 250 | Steel | 150 | 70,000 | 58 | 79 |
| Steel | 200 | Steel | 200 | 70,000 | 58 | 79 |
| Steel | 250 | Steel | 200 | 80,000 | 76 | 103 |
| Steel | 300 | Steel | 200 | 90,000 | 96 | 131 |
| Steel | 250 | Steel | 250 | 90,000 | 96 | 131 |
| Steel | 300 | Steel | 250 | 100,000 | 119 | 162 |
| Steel | 350 | Steel | 250 | 110,000 | 144 | 196 |
| Steel | 300 | Steel | 300 | 110,000 | 144 | 196 |
| Steel | 350 | Steel | 300 | 120,000 | 171 | 233 |
| Steel | 400 | Steel | 300 | 125,000 | 186 | 254 |
| Steel | 350 | Steel | 350 | 130,000 | 201 | 275 |
| Steel | 400 | Steel | 350 | 140,000 | 233 | 318 |
| Steel | 400 | Steel | 400 | 150,000 | 268 | 366 |
| Steel | 150 | Cast iron | | 50,000 | 44 | 60 |
| Steel | 200 | Cast iron | | 70,000 | 87 | 119 |
| Steel | 250 | Cast iron | | 90,000 | 144 | 196 |
| Steel | 150 | Ph. bronze | | 50,000 | 46 | 62 |
| Steel | 200 | Ph. bronze | | 70,000 | 91 | 124 |
| Steel | 250 | Ph. bronze | | 85,000 | 135 | 204 |
| Cast iron | | Cast iron | | 90,000 | 193 | 284 |

Manual of Gear Design, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

TABLE 15-7 VALUES OF LOAD-STRESS FACTOR K FOR HARDENED STEEL

| 10,000,000 Repetitions of Stress | | | | 20,000,000 Repetitions of Stress | | | |
|----------------------------------|---------|------------------|--------------|-----------------------------------|---------|------------------|--------------|
| Brinnell Number | s_c | K (14-1/2°) | K (20°) | Brinnell Number | s_c | K (14-1/2°) | K (20°) |
| 450 | 188,000 | 421 | 575 | 450 | 170,000 | 344 | 470 |
| 500 | 210,000 | 525 | 718 | 500 | 190,000 | 430 | 588 |
| 550 | 233,000 | 647 | 884 | 550 | 210,000 | 525 | 718 |
| 600 | 255,000 | 775 | 1058 | 600 | 230,000 | 630 | 861 |
| 50,000,000 Repetitions of Stress | | | | 100,000,000 Repetitions of Stress | | | |
| 450 | 147,000 | 257 | 351 | 450 | 132,000 | 208 | 284 |
| 500 | 165,000 | 324 | 443 | 500 | 148,000 | 261 | 356 |
| 550 | 182,000 | 394 | 544 | 550 | 163,000 | 316 | 432 |
| 600 | 200,000 | 476 | 651 | 600 | 179,000 | 382 | 522 |

Manual of Gear Design, Vol. II, by Earle Buckingham, The Industrial Press, New York City, with permission.

15-1.10 NONIDEAL CHARACTERISTICS OF GEARS AND GEAR TRAINS

In many control applications, the nonideal characteristics of gears are of extreme importance. In some applications, for example, the choice of gear trains is actually dictated by these characteristics. The subsequent paragraphs cover the following nonideal characteristics: inaccuracies, friction, inertia, backlash, and compliance.

15-1.11 Inaccuracies

Three types of inaccuracies are present to some extent in all gears.⁽⁵⁾ These inaccuracies are:

(a) *Departure of tooth form from the true involute.* The error in action figures given in Table 15-2 indicate the deviation from the true involute form that may be expected in actual production gears. Tooth errors of less

than 0.0002 inch at the pitch line are difficult to achieve, and are obtained only by a high degree of accuracy in the machine on which the gears are cut. Often, however, such highly accurate gears are obtained from a routine production run by careful selection. In general, the more-precise profiles are obtained only at increased cost. Most gears for use in control work can be obtained with tooth-profile errors of less than 0.001 inch, and gears of 20 diametric pitch or finer can be produced with tooth errors of less than 0.0005 inch at little additional cost.

(b) *Errors in tooth spacing.* Tooth-spacing errors are the result of indexing errors, and usually are a result of the inaccuracy of the machine on which the gear is cut. Gears with a maximum tooth-spacing error below 15 seconds of arc are seldom produced, the usual maximum accumulated spacing errors being about 3 minutes of arc.

(c) *Eccentricity of the gear pitch circle.* Eccentricity is produced primarily by errors in holding the gear blank in exact alignment while the teeth are being cut. Eccentricities can be held as low as 0.00025 inch on suitably designed gears, but larger values should be specified as permissible if they are compatible with other design values. It should be remembered that bearing and shaft eccentricities also contribute to the total pitch-line eccentricity of a gear when it is installed in a gear train.

In the majority of control system applications, gear inaccuracies become calibration errors. Two examples of applications where this occurs are: the gears driving the feedback synchros of a gun mount; and the gears driving the synchros of a gun director. Unless extreme accuracy is required, gear inaccuracies in themselves contribute little to control system inaccuracies. However, eccentricities of the gears may necessitate an excessive amount of backlash in the gear train, and this may produce system instability or low system gain. The designer must therefore consider the accuracy requirements of a system when specifying permissible tolerances in gear accuracy.

15-1.12 Friction^(29, 30)

Spur, internal, and bevel gearings usually do not have sufficient friction to warrant consideration. Spiral gears or worms with helix angles below 10°, however, are extremely inefficient because of friction. In general, spiral gears or worms with helix angles below 20° are not reversible; i.e., the worm cannot be rotated by applying torque to the worm gear. In control systems, nonreversible gearing should not be used between the servomotor and the feedback device if the load has any inertia or other source of overdriving torques. When the load tends to overhaul the servomotor, the worm or helical gear locks, and the result is extremely poor servo performance. Improperly designed planetary-gear drives sometime exhibit the same characteristics because high stresses on

the gear teeth produce excessive deformations and consequent high friction.

15-1.13 Inertia

Gearing inertia is often an important portion of the load on a servomotor. The inertia of any gear in a gear train, referred to the driving point, is

$$J_r = \frac{J}{N^2} \quad (15-16)$$

where J is the actual gear inertia, and N is the gear ratio between the reference point and the gear being considered. If the considered gear rotates slower than the reference point, N is greater than unity. If the considered gear rotates faster than the reference point, N is less than unity. The total moment of inertia of the gear train is the sum of all the reflected inertias. The inertia of small-diameter gears is obviously lower than the inertia of larger gears, and the size of the gears in a train can be kept small by mating them with pinions having a minimum number of teeth. In practice, therefore, the total reflected inertia of a gear train can be reduced by using several gear pairs with ratios of 2:1 or less, instead of a single pair with a large ratio. It is often necessary to compromise between reflected inertia and the backlash introduced by a large number of gear meshes.

15-1.14 Backlash

Backlash results in a region of input motion within which no variation of the output can be detected. When placed within a control-system closed loop, this type of "dead zone" may produce system instability, and usually causes the system to oscillate at a low frequency, through an angle of one to five times the backlash angle. Backlash can be reduced by making the driving gear large (to reduce the angle of backlash) or by using gears with small backlash. Backlash can be eliminated by spring loading the gears.

Three methods of eliminating backlash by spring loading are shown in Fig. 15-3. The method shown in Fig. 15-3A uses a spring

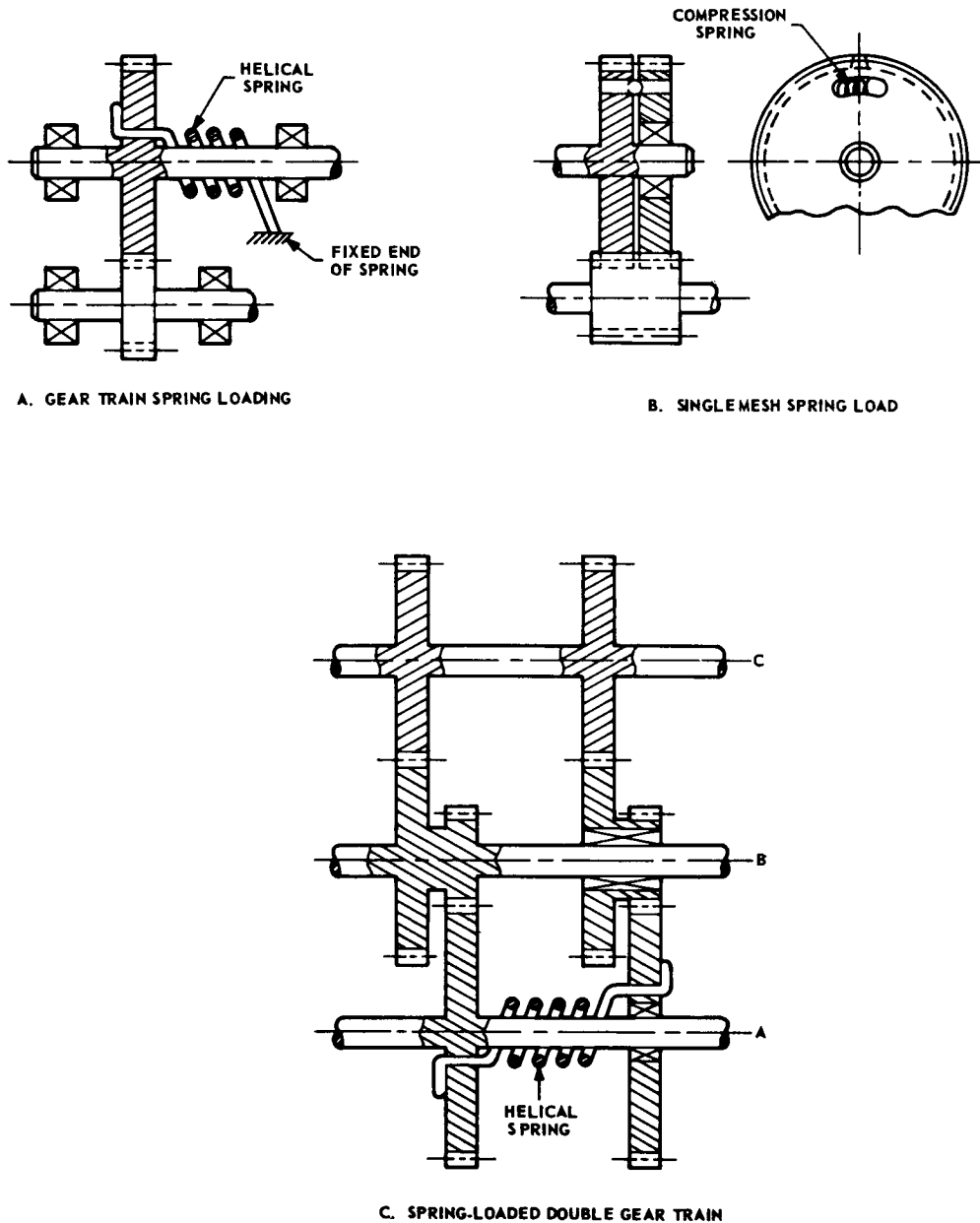


Fig. 15-3 Methods of eliminating backlash.

on the output shaft to exert enough continuous torque to maintain a load on one side of the gear teeth. This method can be used only when there is ample torque available to drive the load, and only when the angle of rotation of the gear is limited to a few revolutions. In Fig. 15-3B, the gear is split in half, forming two identical gears. One gear is attached rigidly to the shaft, and the other is free to rotate on the shaft when driven by expansion of the spring. When the gears are to be meshed with a pinion, the free gear is rotated manually on the shaft to compress the spring. After meshing, the free gear is released and the spring presses the teeth on the two halves of the split gear against opposite sides of the pinion teeth, thus maintaining bidirectional contact between pinion and gear teeth. The method shown in Fig. 15-3C is similar to the split-gear method, except that it can be used to simultaneously load several gear meshes. In this method, the pinion on shaft A is split and spring loaded, thereby loading the entire train. The gear trains shown in Fig. 15-3B and 15-3C are capable of revolving continuously and therefore have a constant loading torque. Any of the three spring-loaded gear-train systems shown in Fig. 15-3 adds to the torque required to drive the gear train because it adds to the load on the bearings and to deformation of the gear teeth.

The optimum amount of spring loading applied to a gear train can sometimes be determined only by trial-and-error. Theoretically, the spring load should be at least equal to the maximum torque the gears are required to transmit. In practice, however, the spring load can often be reduced below this value,

with a resulting decrease in friction in the train. Such decreased loading torques are practicable in applications where the load is primarily inertia, and errors during acceleration are permissible.

15-1.15 Compliance

All gear trains have an effective spring constant as the result of elasticity of the materials. In the presence of heavy inertia or coulomb-friction loads, the compliance (reciprocal of spring constant) of the output gear train of a servomechanism may result in a low system gain, or a highly undesirable resonant frequency. Gear-train compliance comes from the following three sources:

- (a) Elastic deformation of the gear teeth
- (b) Torsion and bending of the shafts
- (c) Deformation of the bearings and housing.

Normally, torsion and bending of the shafts are the only significant factors. The total compliance of a gear train can be computed by adding the compliances of each gear shaft after referring them to a common point. The referred compliance of a shaft is

$$C_r = N^2 C \quad (15-17)$$

where

C = actual compliance of the shaft

N = gear ratio between shaft and reference

C_r = referred compliance

N is greater than 1.0 when the actual shaft rotates slower than the reference shaft, and is less than 1.0 when the actual shaft rotates faster than the reference.

15-2 MECHANICAL DIFFERENTIALS

15-2.1 PURPOSE

Mechanical differentials are used to obtain the algebraic sums of two or more motions. When more than two motions are to be summed, two or more differentials are employed to achieve the required result. The following discussion covers geared differentials and differential linkages.

15-2.2 GEARED DIFFERENTIALS

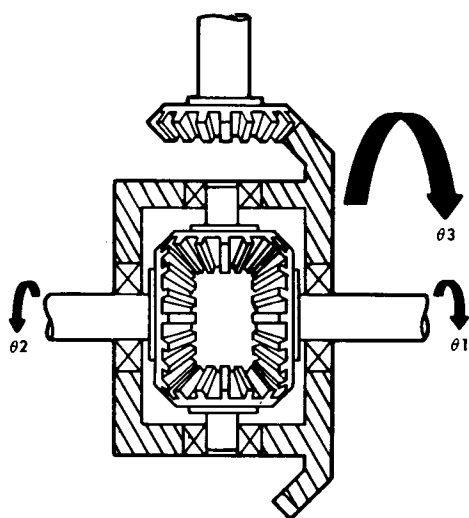
The geared differential is the conventional type of mechanical differential. The most common configurations are shown in Fig. 15-4. Differentials constructed entirely of spur gears are made. However, most differentials incorporate bevel-gear meshes. Bevel-gear differentials are shown in Figs. 15-2A

and B. In both types of differentials, the relation between the rotation of the three shafts is

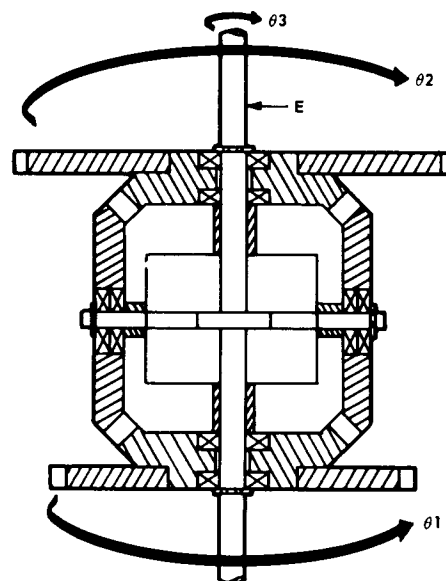
$$2\theta_3 = \theta_1 + \theta_2 \quad (15-18)$$

where θ_1 , θ_2 , and θ_3 are the angles shown in Fig. 15-4.

The number of teeth on meshing gears of a differential is usually selected so that several revolutions of the gears will be made before any particular mesh of gear teeth is repeated. This design reduces and distributes gear wear. Both the good and bad characteristics inherent in any other type of gear train are also existent in geared differentials. Because of the large number of gear meshes in a differential, the amount of backlash is



A. AUTOMOTIVE-TYPE DIFFERENTIALS



B. BEVEL-GEAR DIFFERENTIAL

Fig. 15-4 Geared differentials.

CHAPTER 16

TYPICAL PROCEDURE***16-1 INTRODUCTION**

The design of servo systems to meet a set of specifications can be carried out by either a trial-and-error or an optimization procedure. The method used for a particular design depends to a large extent upon the nature of the specifications and the manner in which they are stated. In many cases, the way the specifications are stated is controlled by the techniques that are most familiar to the specifier.

The trial-and-error design procedure is widely used and has the advantage of being based on well-established and familiar techniques. An optimization procedure, as described in Ch. 8, is tied very closely to the performance index that is being optimized. If the performance index is representative of actual conditions, the optimization procedure is able to determine once and for all whether the specifications are compatible and if so, whether they can be met. No such certainty is possible if a trial-and-error method is used. The major disadvantage of the optimization procedure is that the performance indices which are easily handled by analytical techniques do not correspond too closely to the behavior that is significant to the specifier. Most performance indices that are close approximations to significant behavior are thus far practically impossible to treat by analytical methods, although analog computers can be used in some cases. An

idealization of a practical performance index can be made, however, so that the idealization is amenable to analysis. An example is the use of the integral-square error criterion as an approximation to the peak error criterion for transient inputs. This is justified since systems designed to minimize the integral-square error will probably also minimize the peak error. An example of a practical performance index significant to the specifier and still analytically convenient for the designer is the use of the mean-square error criterion in applications where noise is present at the input or where the input is best represented as a stochastic signal.

Because of the difficulties involved in the use of optimization procedures in practical cases, a compromise is needed. One possibility is to use idealized performance indices in an optimization procedure and then use the resulting optimum system as a guide in the trial-and-error method.

For those cases where the specifications are direct descriptions of the desired closed-loop transfer function (bandwidth, settling time, peak overshoot, etc.), the design procedure is reduced to one of merely synthesizing a transfer function having the desired characteristics. Since the direct specification of the system transfer function is a statement that omits inputs and disturbances, the freedom of the designer is severely limited. The specifier should be sure to specify inputs representative of the actual inputs that the servo will encounter during operation.

**By L.A. Gould, P.E. Smith, Jr., and J.E. Ward
(Par. 16-8 by J.E. Ward)*

16-2 GATHERING OF SPECIFICATIONS

The first step in the design of a servomechanism is to define what the system is to do and how well it is to do it. As pointed out in Ch. 1, a servomechanism may perform any or all of the following functions:

- (a) Bring about a change in the actual value of the output so that it is in conformity with a desired value at all times.
- (b) Minimize the effect of disturbances; i.e., variables other than the desired value of the output.
- (c) Minimize the effect of varying component performance on the output.

The designer must first determine which one, or combination, of these functions is to be achieved.

The degree of excellence to which the servomechanism must carry out the above functions is established by the servo performance specifications. These specifications, furnished by the system designer, are based on a set of system specifications established by the ultimate user of the system.

Servo performance specifications must be given in terms of the desired value of the output of the servomechanism for a given value of the input signal to the servomechanism. The input signal encountered may be one of the following types:

- (a) Aperiodic, noise free
- (b) Aperiodic, with noise
- (c) Periodic, noise free
- (d) Periodic, with noise
- (e) Stochastic, noise free
- (f) Stochastic, with noise

Noise is regarded as any input-signal variation that is not a measure of the information carried by the input. The desired value of the output will be expressed in terms of a signal consistent with the input signal and will therefore be one of the types listed.

Some typical specifications for the six types of input signals listed above are as follows:

- (a) Aperiodic, noise free (step input).

(1) The system error shall neither exceed a specified maximum value at any time nor exceed a specified value after the transient is over; or, alternatively

(2) The integral-square system error shall not exceed a specified value; or, alternatively

(3) The integral of the time-multiplied absolute value of the system error shall not exceed a specified value.

In addition to one of the preceding requirements, it may also be specified that, after the transient due to the step input has died out, the system error shall not exceed a specified amount in the presence of a specified load (disturbance) that occurs at some specified point in the system. Other possible error specifications are listed in Ch. 7.

- (b) Aperiodic, with noise (step input).

(1) The square root of the sum of the squares of the two components of system error shall not exceed a specified value. The first component of the system error is that found for an input step with zero noise; the second component is the value of the output when the input is the noise alone.

Other aperiodic signals commonly considered are ramps, impulse, pulse, and various inputs chosen to simulate target behavior, such as arc-tangent curves or an input expressed as a power series in time t . For inputs of the latter class, it may sometimes be specified that the error coefficients or the error constants, such as the velocity constant or the torque constant, shall have certain maximum or minimum values, respectively. These constants, in a sense, are derived specifications for they actually are chosen with the idea of meeting a certain maximum allowable value of the system error.

(c) Periodic, noise free (only fundamental frequency present).

(1) The output/input frequency response shall be characterized by a specified M_p occurring at a frequency of w_m radians/sec; or, alternatively

TYPICAL PROCEDURE

(2) The output/input magnitude ratio shall be within a band of some specified number of decibels over a specified frequency range, and the phase shift shall not exceed a specified number of degrees over this same frequency range.

(d) *Periodic, with noise.*

(1) The square root of the sum of the squares of the two components of the system error shall not exceed a specified value. The first component of system error is the magnitude of the error due to a sinusoidal input of specified amplitude and frequency. The second component is the rms value of the output when the input is the noise alone.

(e) *Stochastic, noise free.*

(1) The system error shall not exceed a specified rms value when the input autocorrelation function has a given value.

(f) *Stochastic, with noise.*

(1) The system error shall not exceed a specified rms value when the input-signal autocorrelation function, the input-noise autocorrelation function, and the signal-to-noise crosscorrelation function are given.

System performance in response to a given load, disturbance, or combination of disturbances occurring at points different from the input must also be specified. The load or disturbance can be classed as aperiodic, periodic, or stochastic. Typical load specifications, given in terms of the load signal, are as follows:

(a) *Aperiodic load.*

(1) The output shall not deviate from the desired value by more than a specified amount in the presence of a step change in load of a given value; and

(2) The output shall not deviate from the desired value by more than a specified amount for a steady load of specified value; and

(3) The output shall recover to within a specified value of deviation before the elapse of a given time.

(b) *Stochastic load or disturbance.*

(1) The output due to a stochastic load

or disturbance of given autocorrelation function shall not have an rms value in excess of a specified value.

In addition to the specifications on system performance discussed above, the servo-mechanism design is influenced by environmental conditions. The designer must know the allowable values of such physical properties as over-all servo weight and size. Characteristics of the available source of power must also be known. Important among these characteristics are:

(a) *Form*, such as 60-cps or 400-cps alternating current, hydraulic oil at 3000 psi, compressed air at 10 psi, 440-volt direct current, etc.

(b) *Capacity*, such as 115 volts, 15 amperes, 3000 psi at a maximum of 2 ft³/min, etc.

(c) *Regulation*, such as 115 ± 1 volt at 400 ± 2 cps.

The ambient temperature range is significant since it may necessitate the choice of components having special temperature characteristics. Other ambient conditions such as pressure and humidity may be significant. Applicable military specifications for fulfilling environmental conditions must, of course, be followed in the production model. The constraints imposed by safety requirements should also be known; e.g., flammable hydraulic oil may be forbidden in certain applications. Finally, cost may be an important consideration, especially if large numbers of units are to be constructed.

It may often happen that the specifications given to the servo designer are either incomplete, incompatible, or incomprehensible. Incompatibility has been discussed in the beginning of this chapter. Incomplete specifications are those that fail to limit the designer to any particular servo performance. Incomprehensible specifications must be clarified through consultation with the specifier. In the event that the specifier does not have information about the input signal or disturbance, it may be necessary to take sufficient measurements in order to determine an

autocorrelation function. Radar noise may have to be measured and its autocorrelation function found. Typical target paths may have to be decided upon and translated into input-signal variations. In general, designer and specifier have to work together to resolve the difficulties raised by incomplete specifications. As a last resort, where formulation of precise and complete specifications is impossible, estimation based on past experience or on the experience of similar servomechanisms must be used.

As a rule, it is more convenient to carry out the trial design in terms of frequency response as discussed in Chs. 5 and 6. Since the input-output specifications may be in terms of transient behavior, some means of conversion is useful. Once the synthesis has been completed, analytical determination of the exact transient response will have to be carried out. For example, if transient rise

time and output overshoot are specified, the approximate equivalent frequency-response specifications may be deduced through use of the methods outlined in Ch. 7. If the input is given in terms of a specified time function, and if the maximum allowable system error is specified, it may be feasible to determine the various rates of change of input, estimate the error coefficients that result in a total error within specifications, and then proceed with the design to achieve these error coefficients as outlined in Ch. 7.

If the performance specifications do not include a degree of stability in terms of overshoot or frequency response, the designer may arbitrarily pick a value of M_p to be used in the trial-and-error method employing frequency-response techniques. A value of 1.5 decilogs is a reasonable start for many systems. The effect of varying the M_p on the different performance indices should be assessed.

16-3 CHOICE OF TRIAL COMPONENTS*

16-3.1 GENERAL

Once the servo performance specifications have been assembled, a list of suitable power elements, sensing elements, error detectors, and amplifiers can be drawn up. If the load imposed on the servomechanism is known, the size of the output member or power member may be determined. The procedure suggested by Newton⁽¹⁾ may be followed. The method that leads to a choice of motor and gear train (gear ratio) is summarized in this section. Newton's method is applicable to any sort of power element. (See Chapter 14 for descriptions of various sorts of power elements). It is necessary to characterize the power element in terms of its peak available torque, T_{MP} , its maximum allowable velocity,

v_{MP} , and the inertia associated with its output shaft, J_m . Other procedures for choice of output motor can be found in Chestnut and Mayer⁽⁵⁾ and Ahrendt.⁽⁶⁾

16-3.2 DETERMINATION OF MOTOR SIZE

To determine motor size, the following trial-and-error procedure is used:

(a) On the basis of the application specifications, compute the load peak power P_{LP} and the load time constant τ_{LP} by means of the following equations:

$$P_{LP} = \frac{\Delta}{v_{LP}} (J_L a_{LP} + T_{LP}) \quad (16-1)$$

$$\tau_{LP} = \frac{\Delta}{a_{LP}} \quad (16-2)$$

where

P_{LP} = load peak power

*See Appendix on page 16-8 for equation derivations.

TYPICAL PROCEDURE

v_{LP} = load peak velocity

J_L = load inertia

a_{LP} = load peak acceleration

T_{LP} = load peak torque other than inertial

τ_{LP} = load time constant

v_{LP} , T_{LP} , a_{LP} are assumed to act at random, and a consistent set of units are used so that τ_{LP} has a dimension of seconds. One consistent set of units might be the following: v_{LP} in radians/sec, T_{LP} in ft-lb, a_{LP} in radians/sec², J_L in ft-lb-sec², and P_{LP} in ft-lb/sec. Other equally correct sets could be used if desired.

(b) Estimate a preliminary figure for the required motor peak power P_{MP} . (This first guess may be taken to be the load peak power multiplied by a factor equal to one plus a fraction indicated by the designer's experience. A good number is 1.3.)

(c) Referring to data tables for the line of motors under consideration, tentatively choose the motor with the next greater peak power than the estimated figure. Denote this motor peak power by P_{MP-ACT} .

(d) Compute the ratio of the motor time constant τ_{MP} to the load time constant τ_{LP} by use of Eq. (16-3) and the value previously found in Eq. (16-2).

$$\tau_{MP} \triangleq \frac{v_{MP} T_{MP}}{\left(\frac{T_{MP}^2}{J_M}\right)} \triangleq \frac{P_{MP}}{\left(\frac{T_{MP}^2}{J_M}\right)} \quad (16-3)$$

where

τ_{MP} = motor time constant

v_{MP} = motor peak velocity available for the motor chosen in (c)

T_{MP} = motor peak torque available for the motor chosen in (c)

J_M = motor inertia for the motor chosen in (c)

P_{MP} = motor peak power = $v_{MP} T_{MP}$

and a consistent set of units are used so that τ_{MP} has a dimension of seconds.

(e) Find the minimum motor-to-load peak power ratio P_{MP}/P_{LP} for this time-constant ratio by means of the appropriate relation

$$\frac{P_{MP}}{P_{LP}} \geq \frac{1}{1 - \frac{\tau_{MP}}{\tau_{LP}}}, \text{ if } \frac{\tau_{MP}}{\tau_{LP}} \leq \frac{1}{2}$$

or

$$\frac{P_{MP}}{P_{LP}} \geq 4 \frac{\tau_{MP}}{\tau_{LP}}, \text{ if } \frac{\tau_{LP}}{\tau_{MP}} \geq \frac{1}{2}$$

(f) Using this minimum power ratio and the value of P_{LP} from Eq. (16-1), compute the minimum allowable motor peak power P_{MP-MIN} . Find P_{MP-ACT}/P_{MP-MIN} , which is defined as an "over-power" factor F_o . F_o must be equal to, or greater than, unity. If it is not, return to step (c), pick a larger motor, and repeat steps (d) to (f). Continue until a motor with $F_o \geq 1$ is found.

16-3.3 DETERMINATION OF GEAR RATIO

The gear ratio R between motor and load can be found from the following equations, which yield minimum and maximum values of R :

$$\left. \frac{R}{\left(\frac{v_{MP}}{v_{LP}}\right)} \right|_{\min} = \frac{1}{2} \times \frac{1}{\left(\frac{\tau_{MP}}{\tau_{LP}}\right)} \left\{ 1 - \sqrt{1 - 4 \left[\frac{\left(\frac{\tau_{MP}}{\tau_{LP}}\right) - \left(\frac{\tau_{MP}}{\tau_{LP}}\right)^2}{F_o} \right]} \right\} \quad (16-4)$$

$$\text{for } \frac{\tau_{MP}}{\tau_{LP}} \leq \frac{1}{2}$$

$$\left. \frac{R}{\left(\frac{v_{MP}}{v_{LP}} \right)} \right|_{\min} = \frac{1}{2} \times \frac{1}{\left(\frac{\tau_{MP}}{\tau_{LP}} \right)} \left[1 - \sqrt{1 - \frac{1}{F_o}} \right] \quad (16-5)$$

$$\text{for } \frac{\tau_{MP}}{\tau_{LP}} \geq \frac{1}{2}$$

$$\left. \frac{R}{\left(\frac{v_{MP}}{v_{LP}} \right)} \right|_{\max} = 1 \quad \text{for } \frac{\tau_{MP}}{\tau_{LP}} \leq \frac{1 + \sqrt{1 - \frac{1}{F_o}}}{2} \quad (16-6)$$

$$\left. \frac{R}{\left(\frac{v_{MP}}{v_{LP}} \right)} \right|_{\max} = \frac{1}{2} \times \frac{1}{\left(\frac{\tau_{MP}}{\tau_{LP}} \right)} \left[1 + \sqrt{1 - \frac{1}{F_o}} \right] \quad (16-7)$$

$$\text{for } \frac{\tau_{MP}}{\tau_{LP}} \geq \frac{1 + \sqrt{1 - \frac{1}{F_o}}}{2}$$

16-3.4 SELECTION OF OUTPUT MEMBER AND REMAINING ELEMENTS

The proper size motor may be available in either hydraulic or electric form. The choice between the various types of output members must be made on the basis of available power supply and on the relative advantages and disadvantages of the various output members cited in Ch. 14. Although it is possible to find some hydraulic or electric motor to position any given load, it may be found that the weight of the hydraulic motor is substantially less than that of the electric motor in certain cases (see Newton⁽¹⁾). However, the total weight of the servo including amplifiers and auxiliaries must be considered; it is possible that the weight of power supplies may lessen the comparative advantage of the hydraulic motor in some applications. Ready

availability of energy in convenient form elsewhere in a system may vitally influence the choice of which type of servomotor to use. Ease of maintenance and trouble-free operation are factors which must be considered in the light of expected ambient conditions. *Up-to-date* manufacturers' data should be considered for these items as well as in matters of size, weight, and cost of modern component parts. A general comparison of motor types which serves to illustrate some of the significant relative factors appears in Table 16-1. Final choice must be guided by up-to-date manufacturers' information.

Having chosen the output member, the designer must now choose the remaining elements. The set selected must be compatible; e.g., the output member may require force to cause its output to vary. Hence, a source

TYPICAL PROCEDURE

of force (such as a hydraulic amplifier or an electromechanical transducer) must be chosen based on availability and capability of the various types. An amplifier with an output of proper form and capability must then be chosen to drive the force source. The dynamic behavior of the various components chosen for the trial set should be reasonably compatible. For example, if a hydraulic valve and piston are chosen as an output member in order to achieve a servo natural frequency of 30 cps, it would be poor practice to stroke

the valve from a synchro differential having a 2-cps natural frequency. On the other hand, if a servo with a narrow bandwidth is desired, the use of expensive hydraulic components with wide frequency response is not recommended. The measuring and error-sensing device (including the reference) must be chosen with particular attention given to accuracy, since the accuracy of the servomechanism is based on that of its measuring device.

TABLE 16-1 COMPARISON OF SERVOMOTORS

| | D-C Servomotors | A-C Servomotors | Pneumatic Actuators | Hydraulic Motors | Magnetic- Particle Clutches |
|--|--|--------------------|--------------------------------|--|-----------------------------------|
| Efficiency | High | Low | High | High | Low |
| Commutators and brushes | Problem | None | None | None | None |
| Filter required | Electrical Radio Noise Suppression | None | Remove foreign particles | Remove foreign particles | None |
| Flammability | Low | Low | None | High (unless special oils used) | Low |
| Explosion hazard | High | Medium | Low | Low | Medium |
| Coulomb friction, stiction, dead-spot | High | Low | Medium | Medium | Medium |
| Lifetime | Medium | Long | Medium | Medium | Short |

APPENDIX FOR PARAGRAPH 16-3

**DERIVATION OF EQUATIONS PERTAINING TO CHOICE
OF MOTOR AND GEAR TRAIN**

Assumptions

Motor has capability of delivering peak torque T_{MP}

Motor can operate at peak velocity v_{MP}

Motor has an inertia J_M

Load requires a speed-independent torque T_{LP}

Load requires a peak velocity v_{LP}

Load requires a peak acceleration a_{LP}

Load inertia is J_L

Gear-train reduction ratio from motor to load is R .

Units. Any self-consistent set can be used. Two examples:

| Torque | Velocity | Acceleration | Inertia |
|---------------|-------------|--------------------------|------------------------|
| newton-meters | radians/sec | radians/sec ² | kg-meter ² |
| foot-lbs | radians/sec | radians/sec ² | lb-ft-sec ² |

Derivation

Peak motor torque must at least equal total load torque referred to motor; thus:

$$T_{MP} \geq \frac{J_L a_{LP} + T_{LP}}{R} + R J_M a_{LP} \quad (1)$$

Peak motor speed must at least equal required load speed referred to motor; thus:

$$v_{MP} \geq R v_{LP} \quad (2)$$

From Eq. (2), the gear ratio must be as follows:

$$R \leq \frac{v_{MP}}{v_{LP}} \quad (3)$$

Eq. (1) sets upper and lower limits on R . Equation (3) sets an independent upper limit on R . In order for the motor to drive the load, the minimum ratio allowed by Eq. (1) must be less than or equal to the maximum value allowed by Eq. (3). Solve Eq. (1) for minimum value of R ; thus:

$$R_{min} = \frac{T_{MP}}{2J_M a_{LP}} - \sqrt{\frac{T_{MP}^2}{4J_M^2 a_{LP}^2} - \frac{J_L a_{LP} + T_{LP}}{J_M a_{LP}}} \quad (4)$$

TYPICAL PROCEDURE

Combine the limits from Eqs. (3) and (4); thus:

$$\frac{v_{MP}}{v_{LP}} \geq \frac{T_{MP}}{2J_M a_{LP}} \left\{ 1 - \sqrt{1 - \frac{4a_{LP} (J_L a_{LP} + T_{LP})}{T_{MP}^2/J_M}} \right\} \quad (5)$$

Note that Eq. (5) involves motor capabilities and load requirements independent of gear train. Two conditions can be found from Eq. (5). First, the radical must be real; therefore:

$$\frac{T_{MP}^2}{J_M} \geq 4a_{LP} (J_L a_{LP} + T_{LP}) \quad (6)$$

To find another condition, rewrite Eq. (5) as follows:

$$\sqrt{1 - \frac{4a_{LP} (J_L a_{LP} + T_{LP})}{T_{MP}^2/J_M}} \geq 1 - 2 \left\{ \frac{a_{LP}}{v_{LP}} \right\} \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \quad (7)$$

Two cases arise. *Case I.* If

$$\frac{2a_{LP}}{v_{LP}} \times \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \geq 1$$

then the right side of Eq. (7) will be less than zero, and so long as the radical is real, Eq. (7) will be satisfied. Thus, for Case I, no additional constraint is found from Eq. (7).

Case II. If

$$\frac{2a_{LP}}{v_{LP}} \times \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \leq 1$$

when Eq. (7) is satisfied, then Eq. (6) will automatically be satisfied. To find the conditions for satisfying Eq. (7), square both sides. There results:

$$\frac{a_{LP}}{v_{LP}} \cdot \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \left\{ 1 - \frac{a_{LP}}{v_{LP}} \times \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \right\} \geq \frac{a_{LP} J_L a_{LP} + T_{LP}}{T_{MP}^2/J_M} \quad (8)$$

To summarize:

Case I. If

$$\frac{a_{LP}}{v_{LP}} \times \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \leq \frac{1}{2}$$

then

$$(v_{MP} T_{MP}) \left\{ 1 - \left(\frac{a_{LP}}{v_{LP}} \right) \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \right\} \geq v_{LP} (J_L a_{LP} + T_{LP}) \quad (9)$$

Case II. If

$$\frac{a_{LP}}{v_{LP}} \times \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \geq \frac{1}{2} \quad (10)$$

then T_{MP}^2/J_M must be greater than $4a_{LP} (J_L a_{LP} + T_{LP})$.

The requirements for Cases I and II can be rewritten by multiplying the equations respectively by

$$\frac{1}{v_{LP} \left(J_L a_{LP} + T_{PL} \left[1 - \frac{a_{LP}}{v_{LP}} \times \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M} \right] \right)} \quad (11)$$

and

$$\frac{v_{MP} T_{MP}}{v_{LP} (J_L a_{LP} + T_{LP})} \left(\frac{T_{MP}^2}{J_M} \right) \quad (12)$$

Now, define certain characteristic times or time constants. Thus :

$$\tau_{MP} = \frac{v_{MP} T_{MP}}{T_{MP}^2/J_M}$$

$$\tau_{LP} = v_{LP}/a_{LP}$$

Also define peak-load and peak-motor powers as:

$$v_{LP} (J_L a_{LP} + T_{LP}) = P_{LP}$$

$$v_{MP} \times T_{MP} = P_{MP}$$

In terms of these constants, the rewritten Cases I and II become:

Case I.

$$\frac{P_{MP}}{P_{LP}} \geq \frac{1}{1 - \frac{\tau_{MP}}{\tau_{LP}}}, \text{ if } \frac{\tau_{MP}}{\tau_{LP}} \leq \frac{1}{2} \quad (13)$$

Case II.

$$\frac{P_{MP}}{P_{LP}} \geq 4 \frac{\tau_{MP}}{\tau_{LP}}, \text{ if } \frac{\tau_{MP}}{\tau_{LP}} \geq \frac{1}{2} \quad (14)$$

The results are used in the manner described in Par. 16-3.2.

Gear Ratio

The following derivation is for the equations used in the choice of a gear ratio. As mentioned in the chapter, it is always possible to find a motor to drive any load as specified. Then it is only necessary to find a gear ratio. The procedure ensures that once the motor is found, a realizable gear ratio will result.

Multiply Eq. (1) by the factor

$$\frac{\left(\frac{v_{MP}}{P_{MP}} \right) (R)}{\left(\frac{\tau_{MP}}{\tau_{LP}} \right) \left(\frac{v_{MP}}{v_{LP}} \right)}$$

and rearrange Eq. (2). Introduce the time constants.

TYPICAL PROCEDURE

The result is :

$$\left(\frac{1}{\frac{\tau_{MP}}{\tau_{LP}}} \right) \left(\frac{R}{\frac{v_{MP}}{v_{LP}}} \right) \cong \frac{1}{\left(\frac{\tau_{MP}}{\tau_{LP}} \right)} \times \frac{1}{\left(\frac{P_{MP}}{P_{LP}} \right)} + \left(\frac{R}{\frac{v_{MP}}{v_{LP}}} \right)^2 \quad (15)$$

and

$$\frac{R}{v_{MP}/v_{LP}} \leq 1 \quad (16)$$

Consider $\frac{R}{v_{MP}/v_{LP}}$ a nondimensional gear ratio and combine Eqs. (15) and (16) so that the

result is:

$$\frac{1 - \sqrt{1 - \frac{4 \left(\frac{\tau_{MP}}{\tau_{LP}} \right)}{P_{MP}/P_{LP}}}}{2 \left(\frac{\tau_{MP}}{\tau_{LP}} \right)} \leq \left\{ \frac{R}{v_{MP}/v_{LP}} \right\} \leq \begin{cases} 1 \\ \text{or} \\ 1 + \sqrt{\frac{4 \left(\frac{\tau_{MP}}{\tau_{LP}} \right)}{1 - \frac{P_{MP}}{P_{LP}}}} \end{cases} \quad (17)$$

The lower value of the two upper limits sets the upper bound on R . Define a motor over power factor as

$$F_o = \frac{P_{MP-ACT}}{P_{LP}} \times \frac{1}{\left. \frac{P_{MP}}{P_{LP}} \right|_{min}} \quad (18)$$

where $\left. \frac{P_{MP}}{P_{LP}} \right|_{min}$ is the minimum allowable power ratio and $\frac{P_{MP-ACT}}{P_{LP}}$ is the actual power ratio

for the motor chosen. From Eqs. (13) and (14), it is seen that

$$\left. \frac{P_{MP}}{P_{LP}} \right|_{min} = \frac{1}{1 - \frac{\tau_{MP}}{\tau_{LP}}}$$

for the case where

$$\frac{\tau_{MP}}{\tau_{LP}} \leq \frac{1}{2}$$

and

$$\left. \frac{P_{MP}}{P_{LP}} \right|_{min} = 4 \frac{\tau_{MP}}{\tau_{LP}}$$

POWER ELEMENTS AND SYSTEM DESIGN

for the case where

$$(\tau_{MP}/\tau_{LP}) \geq \frac{1}{2}$$

Solve Eq. (18) for the actual value of P_{MP}/P_{LP} ; thus:

$$\frac{P_{MP-ACT}}{P_{LP}} = F_o \frac{P_{MP}}{P_{LP}} \Big|_{min}$$

Substitute the values for $\frac{P_{MP}}{P_{LP}}$ into Eq. (19) to find:

$$\frac{P_{MP-ACT}}{P_{LP}} = \frac{F_o}{1 - \frac{\tau_{MP}}{\tau_{LP}}}, \text{ if } \frac{\tau_{MP}}{\tau_{LP}} \leq \frac{1}{2}$$

and

$$\frac{P_{MP-ACT}}{P_{LP}} = 4 F_o \frac{\tau_{MP}}{\tau_{LP}}, \text{ if } \frac{\tau_{MP}}{\tau_{LP}} \geq \frac{1}{2} \quad (20)$$

Substitute Eq. (20) into Eq. (17). The two minimum values of $R/(v_{MP}/v_{LP})$ are found directly from the left-hand side of Eq. (17). The maximum values of $R/(v_{MP}/v_{LP})$ are found from either the upper or lower inequality of the right-hand side of Eq. (17). The condition as to whether the upper or lower inequality of Eq. (17) holds depends upon which of the two is the greater. These results all appear as Eqs. (16-4) through (16-7).

CHAPTER 17**REPRESENTATIVE DESIGNS*****17-1 INTRODUCTION**

The examples of servomechanism systems described in this chapter have been selected from Army equipments which now carry an unclassified designation. The objectives of presenting these designs is to illustrate the analytical procedures discussed in the earlier chapters and to show how various components may be integrated to meet a set of over-all requirements. It should be borne in

mind that each new requirement necessitates, in general, a fresh approach to system synthesis in order to achieve best results, and that no single type of system organization can be regarded as the best design for all purposes. The user of the handbook is therefore cautioned against applying the representative designs in this chapter to new situations.

17-2 A SERVO SYSTEM FOR A TRACKING-RADAR ANTENNA**17-2.1 GENERAL**

The M33 Antiaircraft Fire-Control System employs electrical servos for driving the tracking-radar antenna in azimuth and elevation during the tracking of a target. A functional block diagram showing the interconnection of the units is drawn in Fig. 17-1. Only the elevation-control channel is illustrated since both channels are identical with the exception that the azimuth power servo amplifier drives four motors whereas the elevation servo amplifier drives a single motor. The units shown are those used in the automatic-tracking mode.

17-2.2 PURPOSE

The primary purpose of the azimuth and elevation servos is to steer the tracking-radar antenna so that an airborne target may be tracked automatically in two co-

ordinates as it moves through space. Another purpose of the servos is to provide for rapid slewing of the tracking antenna to the angular position of the associated acquisition-radar antenna. Still another purpose is to provide alignment (in elevation) of the optical system with the tracking-radar antenna. The two servos also supply angular-position data to a computer and to remote indicators.

17-2.3 OPERATION

An 1800-rpm conical-scan drive motor attached to an off-axis antenna feed (see Fig. 17-1) spins the beam in space and, in so doing, places 30-cps amplitude modulation on the chain of echo signals received from a target. After amplification and detection by the radar receiver, the 30-cps amplitude-modulated video echo pulses are applied to a pulse demodulator where the modulation signal is recovered from the video pulses. The 30-cps signal is then applied to an elevation-angle

**By J. F. Reintjes and L. A. Gould*

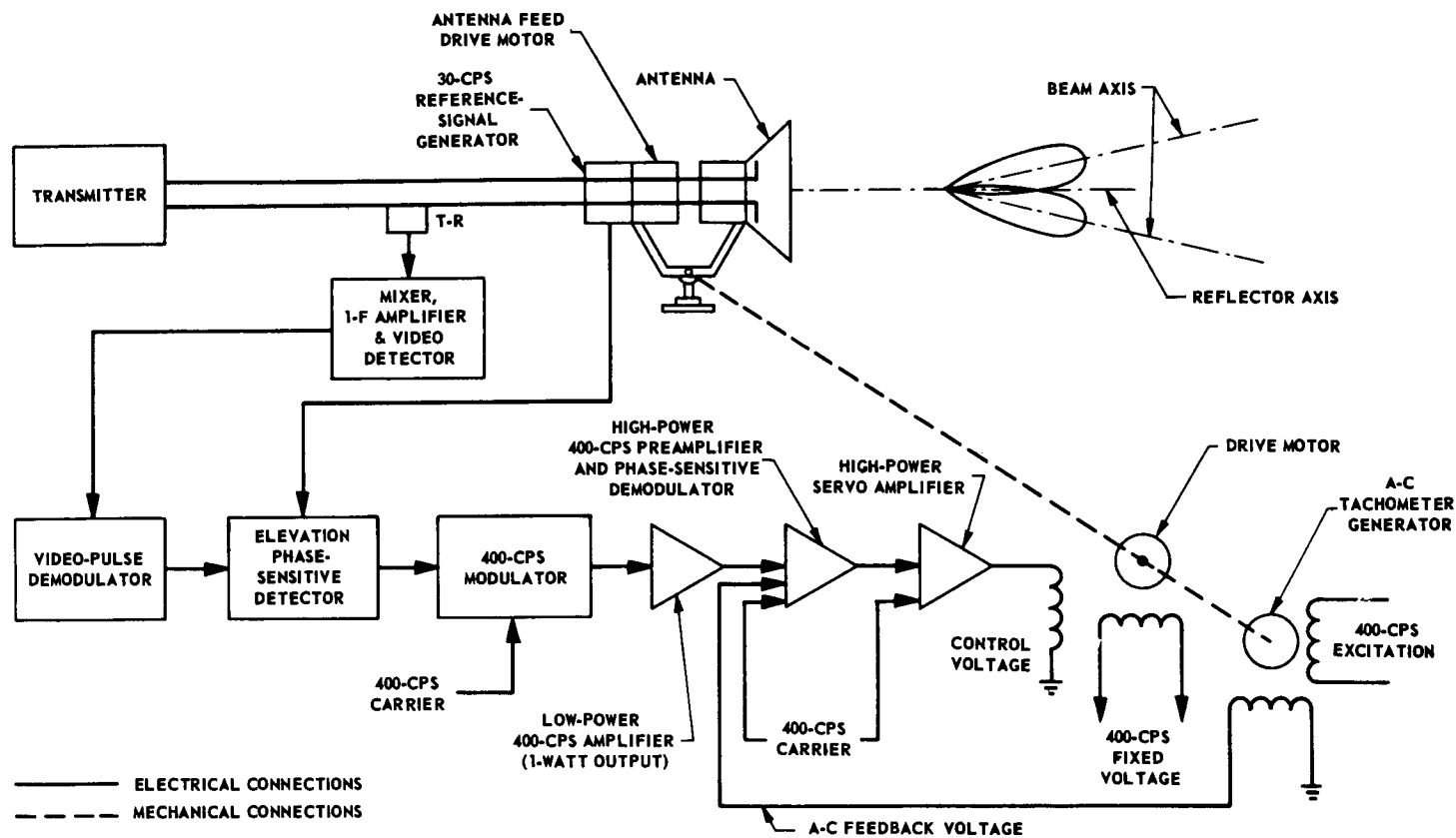


Fig. 17-1 Simplified functional block diagram of servo system for controlling M33 tracking-radar antenna in elevation.

phase-sensitive detector (see Par. 12-4) together with a 30-cps signal from a reference-signal generator coupled to the conical-scan drive motor at the antenna. The reference signal serves as a phase reference for the elevation angle of the antenna-beam axis measured with respect to the parabolic-reflector axis. If the antenna is on target, no error signal is generated in the phase-sensitive detector. If the antenna is off target in elevation, there is generated an error signal whose magnitude is a measure of the amount the antenna-beam axis is off target and whose polarity with respect to the elevation reference signal depends upon whether the beam axis is above or below target. Hence, as the antenna tracks a moving target, the output of the phase-sensitive detector is a slowly varying positive or negative voltage, the magnitude and polarity of which are a measure of the elevation tracking error.

The error signal from the phase-sensitive detector modulates a 400-cps carrier voltage, and the modulated carrier is then amplified in a low-power electronic servo amplifier. This amplifier consists of two stages of voltage amplification, a phase inverter, and a push-pull output stage (see Par. 13-1). The phase inverter provides two signals, 180° out of phase, for driving the push-pull output stage. The output of the low-power servo amplifier is then applied to the input of a 400-cps high-power servo preamplifier. Also applied to the preamplifier input is a 400-cps feedback voltage from an a-c tachometer generator (see Par. 11-5) that is coupled to the elevation drive shaft of the antenna. An a-c velocity feedback voltage is thus provided in the servo system. Because the high-power servo-amplifier stage requires a push-pull d-c voltage input signal, the preamplifier also includes a demodulator. The input signal to the preamplifier is therefore a modulated 400-cps carrier signal and the output is the modulation envelope.

The high-power servo amplifier controls the 400-cps 2-phase elevation drive motor and consists of a push-pull d-c amplifier driving a 400-cps magnetic modulator (see Par. 12-3).

The output of the magnetic modulator is a 400-cps voltage whose amplitude is proportional to the difference between the output currents in each tube of the push-pull amplifier and whose phase is equal or opposite to that of the 400-cps excitation voltage of the magnetic modulator, depending upon which of the two tube currents is greater. The output signal is applied to the control winding of the 2-phase motor.

17-2.4 OPERATIONAL BLOCK DIAGRAM

The operational block diagram of the servo system for controlling the tracking-radar antenna in elevation is shown in Fig. 17-2. In this figure

- r = line-of-sight elevation angle
- c = antenna elevation angle
- e = actuating signal of servo system
- n = radar noise
- θ_m = angle of servomotor shaft
- b = tachometer-generator feedback signal
- u = load-torque disturbance at antenna reflector
- $G_1(s)$ = transfer function of radar set, elevation phase-sensitive detector, 400-cps modulator, and low-power 400-cps amplifier
- $G_2(s)$ = transfer function of high-power 400-cps preamplifier and phase-sensitive demodulator, high-power servo amplifier, and servomotor control field
- $G_3(s)$ = transfer function of servomotor
- $G_4(s)$ = transfer function of antenna mechanical assembly
- $H_1(s)$ = transfer function of tachometer generator
- $H_2(s)$ = transfer function due to imperfect coupling between antenna reflector and servomotor shaft

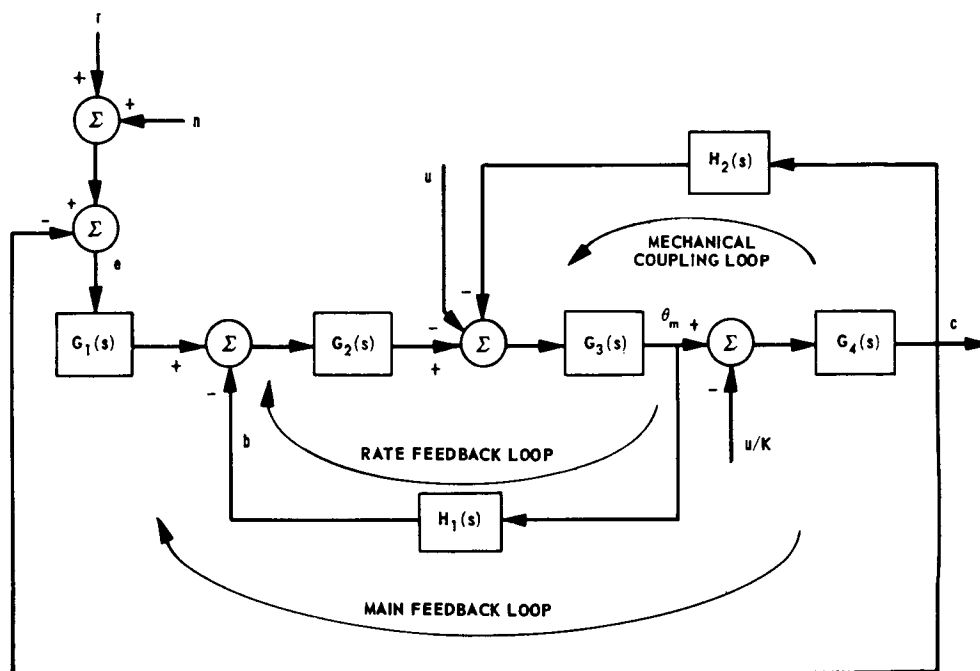


Fig. 17-2 Operational block diagram of servo system for controlling M33 tracking-radar antenna in elevation.

K = stiffness of antenna structure

The rate feedback supplied by the a-c tachometer generator $[H_1(s)]$ is imperfect because the antenna structure is not absolutely rigid. The rate feedback therefore tends to improve the performance of the servomotor but is not completely effective in reducing the mechanical resonance of the antenna structure. The dynamic behavior of the servomotor and the antenna structure cannot be separated in the system, as indicated by the mechanical coupling loop. However, if the resonant frequency of the antenna structure is very high, only the inertia of the structure is important; the rate feedback is then a direct measure of the elevation angular velocity of the antenna reflector.

17-2.5 NOISE

The presence of radar noise in the line-of-sight signal prevents the actuating signal e

from being a true measure of the system error, $y_e = r - c$. In order to minimize the effect of noise, the bandwidth of the system has been kept at a low value so that the servo-mechanism acts as a low-pass filter.

17-2.6 DESIGN CHARACTERISTICS

The following is a list of characteristics of the elevation servo system :

| | |
|----------------------------|----------------------------|
| Maximum power output | 120 w |
| Maximum velocity | 500 mils/sec |
| Maximum acceleration | 1000 mils/sec ² |
| Maximum static error | 0.04 mil |
| Main power requirements | 360 va |
| Control power requirements | 360 va |
| Resonant frequency | 1 cps |
| M_p | 1.5 dg |

17-3 A POWER CONTROL SYSTEM FOR THE M-38 FIRE-CONTROL SYSTEM

17-3.1 GENERAL

The M-38 Fire-Control System[†] is a mobile 75mm antiaircraft weapon designated as the *Skysweeper*. The power control system for this weapon (see Fig. 17-3) is an electrically controlled hydromechanical drive mounted on the carriage of the gun mount, and serves to position the gun in azimuth and elevation in accordance with command signals from the fire-control computer. The azimuth and elevation servo systems are essentially identical.

17-3.2 OPERATION

The fire-control computer provides either present or future target position data, depending upon the mode of computer operation. These data are compared with positional data from the gun. If the desired and actual positions do not correspond, there is generated an error-signal voltage whose amplitude is proportional to the difference between the desired and actual gun positions and whose phase is a measure of whether the gun lags or leads the desired position. The error signal is then amplified in a servo amplifier to a power level sufficient to drive a stroke-control servomotor. This servomotor, in turn, is geared and linked mechanically to the follow-valve control (pilot valve) of a power piston that regulates the oil delivery rate of a hydraulic pump and therefore the output velocity of the hydraulic motor connected to the output gearing.

As indicated in Fig. 17-3, voltages from the computer are supplied to synchro control transformers as 1:1 coarse data and 16:1 fine data. The rotors of the synchros are connected mechanically to the output gearing. The resultant error-signal voltages, after being combined with secondary signals to be

discussed below, are amplified and applied to the power amplifier. For large error-signal voltages, the coarse error signal predominates and serves to swing the gun approximately on target. For smaller angular errors, the fine error signal predominates and serves to keep the gun precisely on target.

The coarse synchro control transformer rotor is geared to make one revolution for each revolution of the gun, whereas the fine synchro control transformer rotor makes sixteen revolutions for each gun revolution. As a result, the envelope of the coarse synchro signal undergoes two phase reversals for each revolution of the gun, while the envelope of the fine synchro signal undergoes thirty-two phase reversals. At the 3200-mil azimuth gun position, therefore, the coarse and fine voltage envelopes are exactly 180° out of phase. Because of this 180° phase relationship, it is possible to synchronize the gun exactly 180° away from its correct position.

To avoid this false synchronization, a small a-c synchronizing voltage, called an *antistick-off* voltage, is placed in series with the rotor voltage of the coarse synchro control transformer. By proper adjustment of the phase angle and amplitude of this auxiliary voltage, it is possible to bring the coarse and fine synchro voltage envelopes into phase agreement in the region of possible false synchronization (see Par. 11-3).

The error signals from the coarse and fine synchro control transformers are in the form of sinusoidal voltages. The error voltage from the fine synchro control transformer is passed through an amplitude limiter so as to limit the maximum error it can correct to approximately 25 mils. The coarse error-voltage waveform, on the other hand, is reshaped so that its amplitude is essentially zero for signals corresponding to an error of 25 mils or less. Hence, the coarse error signal effectively controls circuit operation for tracking

[†]Formerly called the T-38 Fire-Control System.

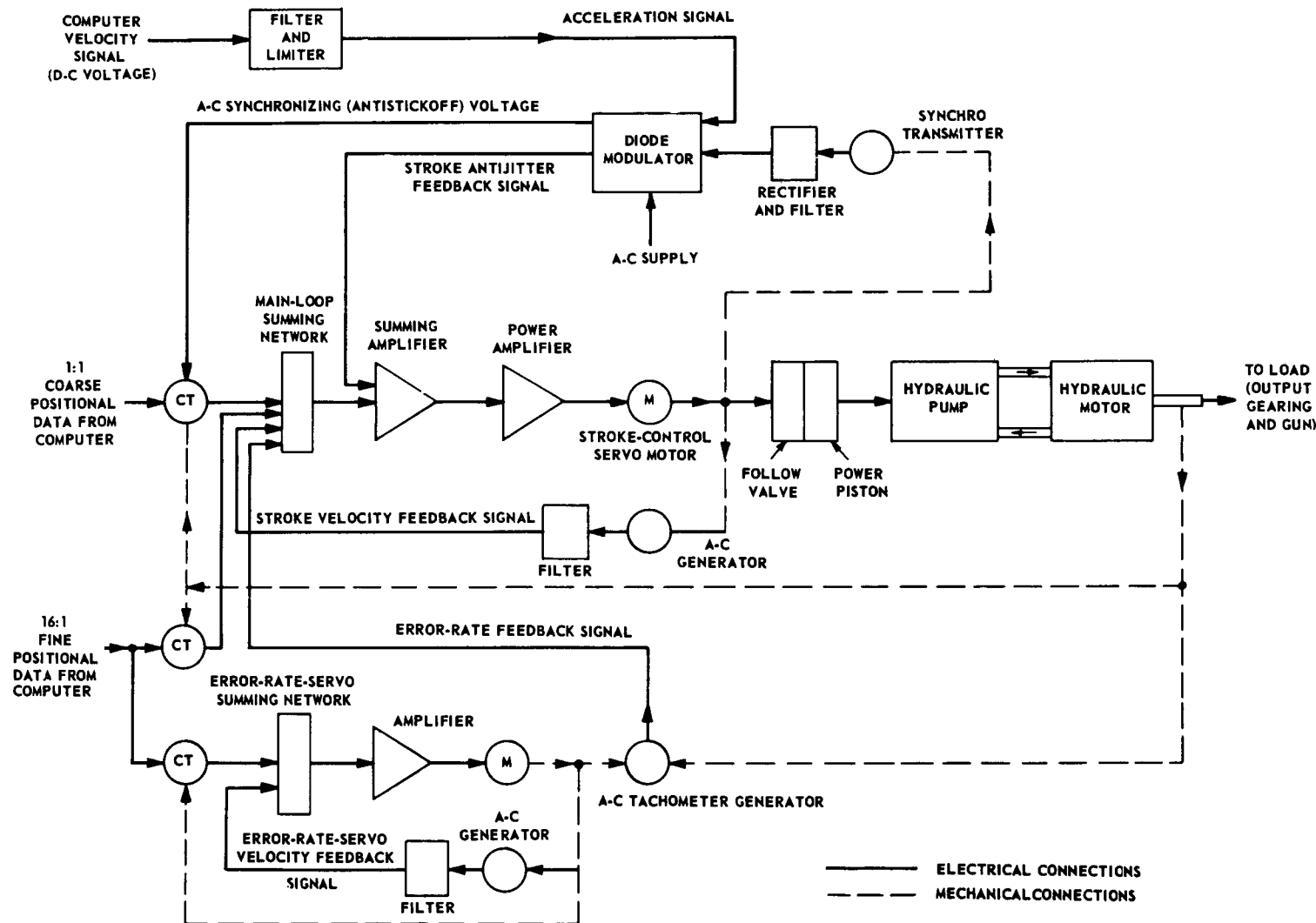


Fig. 17-3 Power control system for M-38 Fire-Control System.

REPRESENTATIVE DESIGNS

errors greater than 25 mils and the fine error signal controls circuit operation for errors less than 25 mils.

The summing amplifier in Fig. 17-3 is similar to that described in Par. 13-1; the power amplifier consists of a phase inverter and a push-pull amplifier. The output of the latter stage drives the control winding of a 2-phase stroke-control servomotor, the direction and speed of motor rotation being dependent upon the amplitude and phase of the amplifier output signal. The servomotor controls the position of a follow valve in a hydraulic servo cylinder (see Par. 14-4) and thus regulates the volume and direction of fluid delivery to the fixed-displacement hydraulic motor geared to the gun. Controlling fluid delivery controls the speed and direction of the hydraulic motor and hence the speed and direction of gun travel.

The coarse and fine positional-signal-control channel described above contains several internal feedback loops. In addition, an auxiliary error-rate servo with its own error and feedback voltages, mixing circuit, amplifier, and servomotor is included in the over-all power control system. To reduce overshoot and to eliminate hunting of the stroke-control servomotor, a stroke velocity signal is fed back to the summing amplifier through a phase-correcting network. The feedback voltage is obtained from an a-c eddy-current generator attached directly to the shaft of the stroke-control servomotor. The amplitude of the generator output voltage is proportional to the speed of shaft rotation of the servomotor, and the phase of the voltage is dependent upon the direction of rotation.

A second feedback path from stroke-control servomotor to summing amplifier serves to suppress small oscillations of the servomotor output shaft that would otherwise be transmitted to the gun. The stroke antijitter voltage is obtained from a modified synchro control transmitter that is friction-coupled to the stroke-control servomotor shaft. The transmitter rotor is excited from a potentiometer connected to a 6.3-volt a-c source. The potentiometer is adjusted to provide the

minimum voltage consistent with the elimination of gun jitter. The output of the transmitter is taken from two series-connected stator windings, and thus the magnitude of the output signal is a function of the displacement of the rotor. The rotor is limited by means of stops to $\pm 30^\circ$ displacement. The output voltage of the transmitter is then rectified, filtered, and differentiated in a rate circuit. All steady voltages are therefore blocked, but transient signals are passed on to the diode modulator circuit.

The diode modulator circuit supplies the summing amplifier with square-wave signals that are in phase with the coarse and fine error signals and that vary in amplitude in proportion to the sum of the stroke antijitter signal and an acceleration signal. The latter signal provides a means of anticipating sudden changes in speed or direction of the gun and is obtained from a velocity signal in the computer. The velocity signal is supplied as a d-c voltage from a permanent-magnet generator. After filtering to remove commutator ripple, the velocity signal is differentiated in order to obtain a voltage that is proportional to acceleration. The acceleration signal is then combined with the stroke antijitter signal as indicated above.

The error-rate servo in Fig. 17-3 receives its input signals from the 16:1 speed transmitter in the computer. These signals are amplified to drive the rotor of an a-c tachometer generator, the stator of which is driven by the gun gearing. The error voltage thus generated is applied, after being phase-synchronized and mixed with the other feedback signals, to the main summing amplifier. As long as the gun turns in synchronism with the fine-speed data from the computer, there is no relative motion between rotor and stator of the a-c tachometer generator and no signal voltage is generated. Should the gun begin to lead or lag the point of synchronization, the phase and magnitude of the error-rate signal will accelerate or decelerate the gun and thus bring the gun velocity into agreement with the velocity of the computer synchro-transmitter rotor.

17-3.3 OPERATIONAL BLOCK DIAGRAM

The detailed operation block diagram of the fine-speed portion of the power control system for the M-38 Fire-Control System is shown in Fig. 17-4. The nomenclature for this diagram is presented in Table 17-1. If the dynamic lags (G_8 , G_9 , G_{10}) associated with the motors and their driving members are neglected, and only low-frequency behavior is considered, the block diagram can be reduced to the form shown in Fig. 17-5, where

ω_1 = magnitude crossover frequency of error-rate servo

ω_2 = resonant frequency of main control loop

T_1 = gain factor associated with error-rate stabilization

T_2 = gain factor associated with anticipatory acceleration signal

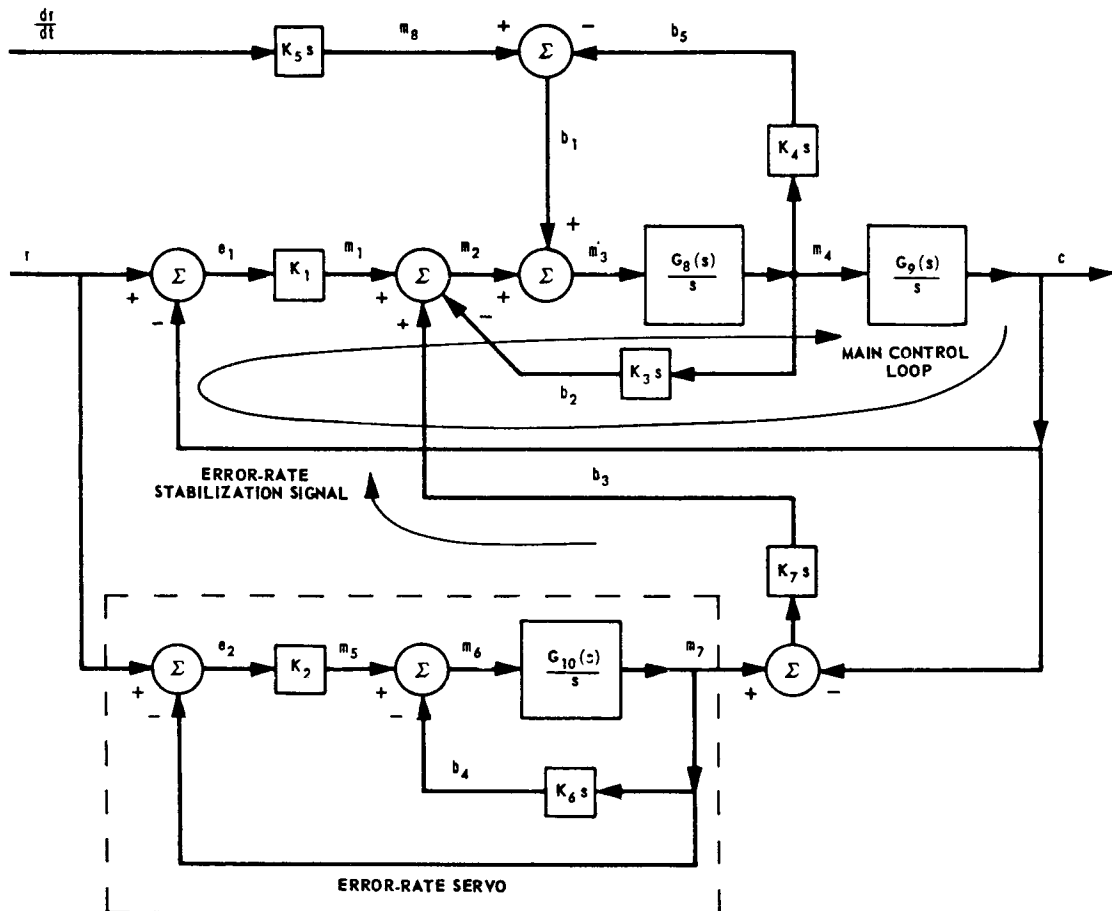


Fig. 17-4 Operational block diagram of fine-speed portion of power control system for M-38 Fire-Control System.

REPRESENTATIVE DESIGNS

It is evident that the main control loop of this system is inherently unstable because of the presence of two integrations, one due to the stroke-control servomotor and one due to the hydraulic motor. The error-rate stabilization signal is therefore introduced to supply the damping that is necessary for stable operation of the main control loop. If the magnitude crossover frequency ω_1 of the error-rate servo is high compared with the resonant frequency ω_2 of the main control loop, the diagram in Fig. 17-5 can be reduced to the form shown in Fig. 17-6. In this form, it is evident that the gain factor T_1 associated with the error-rate stabilization directly controls the damping of the main control loop. Furthermore, the addition of the acceleration signal raises the possibility of cancelling the transfer function of the main control loop (by setting $T_2 = 1/\omega_2^2$) and thus producing an extremely rapid response for the system. Because of the approximations made in the successive block diagram reductions, such a cancellation will not actually occur, but the response of the system can be effectively speeded up. This improvement, of course, is limited by the dynamic lags that have been neglected and by the restriction of linear operation that is implicit in the analysis.

17-3.4 DESIGN CHARACTERISTICS

The over-all azimuth servo characteristics are as follows:

| | |
|-------------------------|----------------------------|
| Maximum power output | 5 hp |
| Maximum velocity | 1050 mils/sec |
| Maximum acceleration | 2000 mils/sec ² |
| Maximum static error | 0.75 mil |
| Main power requirements | 12 kw |

| | |
|----------------------------|---------------|
| Control power requirements | 200 w |
| Resonant frequency | 0.5 and 3 cps |

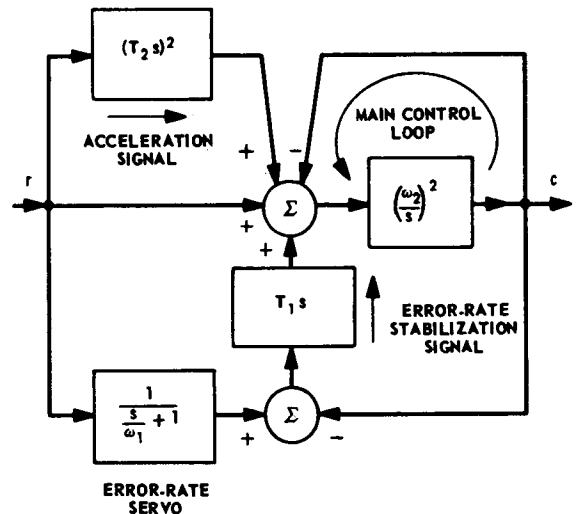


Fig. 17-5 Simplified block diagram of power control system for M-38 Fire-Control System.

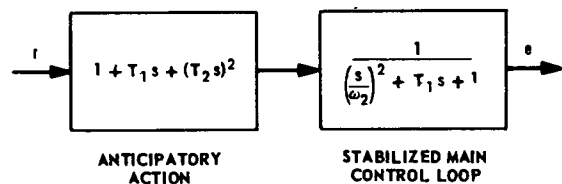


Fig. 17-6 Final block diagram of power control system for M-38 Fire-Control System.

POWER ELEMENTS AND SYSTEM DESIGN

TABLE 17-1 NOMENCLATURE FOR M-38 POWER CONTROL SYSTEM

| | |
|-----------------|---|
| r | = fire-control-computer target position data (reference input) |
| $\frac{dr}{dt}$ | = fire-control-computer velocity data |
| c | = gun position (controlled variable) |
| e_1 | = difference between reference input r and gun position c (main-loop actuating signal) |
| e_2 | = difference between reference input r and error-rate-servo output position m_7 (error-rate-servo actuating signal) |
| m_1 | = output signal of main-loop fine-speed control transformer |
| m_2 | = output signal of main-loop summing network |
| m_3 | = sum of stroke antijitter feedback signal b_1 and output signal m_2 of main-loop summing network |
| m_4 | = stroke-control servomotor position |
| m_5 | = output signal of error-rate-servo control transformer |
| m_6 | = output signal of error-rate-servo summing network |
| m_7 | = error-rate-servo output position |
| m_8 | = acceleration signal |
| b_1 | = stroke antijitter feedback signal |
| b_2 | = stroke velocity feedback signal |
| b_3 | = error-rate feedback signal |
| b_4 | = error-rate-servo velocity feedback signal |
| b_5 | = smoothed and differentiated synchro-transmitter signal |
| K_1 | = main-loop control-transformer sensitivity |
| K_2 | = error-rate-servo control-transformer sensitivity |
| K_3 | = gain factor relating stroke velocity feedback signal b_2 to stroke-control servomotor velocity dm_4/dt |
| K_4 | = gain factor relating differentiated synchro-transmitter signal b_5 to stroke-control servomotor velocity dm_4/dt |
| K_5 | = gain factor relating acceleration signal m_8 to rate of change $d/dt(dr/dt)$ of computer velocity data |
| K_6 | = gain factor relating error-rate-servo velocity feedback signal b_4 to error-rate-servo output velocity dm_7/dt |
| K_7 | = gain factor relating error-rate feedback signal b_3 to difference between error-rate-servo output velocity dm_7/dt and gun velocity dc/dt |
| $G_8(s)$ | = transfer function relating stroke-control servomotor velocity dm_4/dt to sum of stroke antijitter feedback signal b_1 and output m_2 of main-loop summing network |
| $G_9(s)$ | = transfer function relating gun velocity dc/dt to stroke-control servomotor position m_4 |
| $G_{10}(s)$ | = transfer function relating error-rate-servo output velocity dm_7/dt to output m_6 of error-rate-servo summing network |

REPRESENTATIVE DESIGNS

BIBLIOGRAPHY

- 1 *Traverse Drive Mechanism*, SA 2482, Ordnance Division, Minneapolis-Honeywell Regulator Company, Minneapolis, Minn., November, 1954.
- 2 O. Wilsker, *Cant Corrector Servomechanism for Range Finder*, T42, ORDTX-10 Technical Memorandum, Report No. M-52-8-1, Fire Control Instrument Group, Frankford Arsenal, Philadelphia, Pa., July 1, 1952.
- 3 M. D. Swartz, *Fire Control for Tanks; Instrument Servos for Tank Applications*, ORDTX-10 Technical Notes, Report No. TN-1011, Fire Control Instrument Group, Frankford Arsenal, Philadelphia, Pa., January 16, 1953.
- 4 H. A. Smith, *Tabulation of Data for Servo Systems in Existing Antiaircraft Material*, ORDTX-10 Technical Notes, Report No. TN-1051, Fire Control Instrument Group, Frankford Arsenal, Philadelphia, Pa., October 15, 1954.
- 5 *Second Antiaircraft Fire Control Working Conference* (notes compiled by F. Q. Barnett and A. M. Paup), Report No. 932, Ballistic Research Laboratories, Aberdeen Proving Ground, Md., May, 1955 (secret).
- 6 *Notes on Materiel — 75 MM AA Gun Mount T19 Skysweeper, Vol. II, Computer T27, Director T41* (prepared by Sperry Gyroscope Company), Dept. of the Army, FCDD-248, Sperry Publication No. 14-8092, May 25, 1948 (secret).
- 7 *Notes on Development Type Materiel — Fire Control System, AA, T33, Contract W30-069-ORD-4490, Vol. IV, Maintenance Instructions* (prepared by Western Electric Company), Dept. of the Army, FCDD-261, May 1, 1950.
- 8 *Notes on Development Type Materiel — Fire Control System, AA, T33, Contract W30-069-ORD-4490, Vol. II, Wiring Diagrams* (prepared by Western Electric Company), Dept. of the Army, FCDD-261, May 1, 1950.
- 9 *Notes on Development Type Materiel — Fire Control System, AA, T33, Contract W30-069-ORD-4490, Vol. III, Schematic Diagrams* (prepared by Western Electric Company), Dept. of the Army, FCDD-261, May 1, 1950.
- 10 *Notes on Materiel — Fire Control System T38 for 75-MM Gun T83E1 and 75-MM AA Gun Mount T69, Vol. 1, Skysweeper System and Power Control T21E1* (prepared by Sperry Gyroscope Company), Dept. of the Army, FCDD-263, Sperry Publication No. 14-8093, February 1, 1951.
- 11 *Notes on Materiel — Fire Control System T38 for 75-MM Gun T83E1 and 75-MM AA Gun Mount T69, Vol. III, Computer T27E2, etc.* (prepared by Sperry Gyroscope Company), Dept. of the Army, FCDD-263, Sperry Publication No. 14-8093, February 1, 1951.
- 12 *Notes on Development Type Materiel — Ranging and Sighting Equipment for 76-MM Gun Tank T41E1, 90-MM Gun Tanks T42 and M47, and 120-MM Gun Tank T43*, Ordnance Corps, FCDD-274, April, 1952.
- 13 *Notes on Materiel — Short Range AA Weapon (Stinger), Vol. I, Description, Theory, and Operation* (prepared by

POWER ELEMENTS AND SYSTEM DESIGN

Sperry Gyroscope Company), Dept. of the Army, FCDD-291, Sperry Publication No. 14-8094, November 1, 1953 (confidential).

14 *Notes on Materiel — Short Range AA Weapon (Stinger), Vol. II, Maintenance Instructions* (prepared by Sperry Gyroscope Company), Dept. of the Army, FCDD-291, Sperry Publication No. 14-8094, November 1, 1953 (confidential).

15 *Notes on Materiel — Short Range AA*

Weapon (Stinger), Vol. III, Schematics, Servo Loop Diagrams, and Wiring Lists (prepared by Sperry Gyroscope Company), Dept. of the Army, FCDD-291, Sperry Publication No. 14-8094, November 1, 1953.

16 *Notes on Development Type Materiel — Primary Fire Control System, Tank, 90 MM Gun, M48* (prepared by Chrysler Corporation), Dept. of the Army, FCDD-305, Ordnance Project No. TT2-760, July 10, 1955.

CHAPTER 18

AUXILIARIES ASSOCIATED WITH SERVOMECHANISMS*

18-1 AUXILIARY PUMPS

18-1.1 PURPOSE

In a servo system incorporating a hydraulic transmission, hydraulic fluid (oil) lost by leakage from the main hydraulic pump and motor must be replaced, otherwise the system eventually runs dry and ceases to function. Automatic, continuous replacement of lost oil is accomplished by means of auxiliary pumps operating in conjunction with regulating devices. When it is used for this purpose, an auxiliary pump is called a replenishing pump, a supercharging pump, or a make-up pump. The replenishing pressure is usually below 100 psi.

Many hydraulic-transmission servo systems also include hydraulic amplifiers, which control the stroke of the main hydraulic pump. Hydraulic amplifiers require oil at a pressure that may be as high as 1000 psi. Because of high pressure in large servo systems of this type, two auxiliary pumps may be used: one to feed the amplifier and one to feed the replenishing system. In smaller servo systems of this type, one auxiliary pump may perform both functions. In a servo system utilizing a valve-controlled ram or hydraulic motor, the pump used to feed the valve is often considered a subsidiary device because the pump is not in the closed-loop control system.

*By J.O. Silvey

18-1.2 TYPES OF AUXILIARY PUMPS

Any type of pump that supplies the necessary pressure and rate of flow is usually adequate for auxiliary service in a servo system. The type generally used in practice is the fixed-displacement pump driven at a constant speed. The fluid pressure is controlled by means of automatic pressure regulators or unloading valves and accumulators. Occasionally, variable-delivery piston-type pumps are used to maintain a constant pressure at all flow rates between zero and their rated maximum. When used for this purpose, however, the pump is considered as merely one component of a pressure-control system. Variable-delivery pumps are discussed in Par. 13-6. The fixed-displacement pump types discussed in the following paragraphs are:

- (a) Gear pumps
- (b) Vane pumps
- (c) Piston pumps

18-1.3 Gear Pumps

A gear pump comprises two meshed gears enclosed by a close-fitting housing equipped with an inlet (suction) port and a discharge port (Fig. 18-1). One gear is shaft-driven by an external motor. Oil is carried from the inlet port to the discharge port in the spaces between the gear teeth, and is forced from these spaces where the gears mesh.

Usually, spur or helical gears are used in gear pumps. Gear materials range from cast

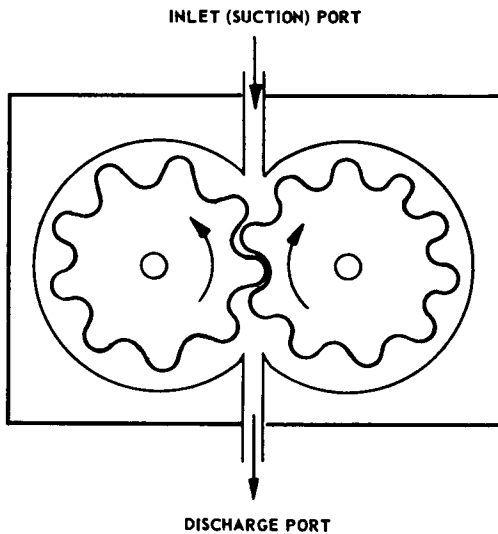


Fig. 18-1 Gear pump.

iron through hardened alloy steel. The material used depends upon the service expected from the pump.

Capacities of gear pumps normally used as replenishing pumps range from 1 to 60 gpm (gallons per minute). Rated output pressures range from 15 psi for replenishing service to 2000 psi for intermittent heavy duty.

Oil leakage from the high-pressure (discharge) side of the gears to the low-pressure (inlet) side is held to a minimum by the close fit of the housing along the sides of the gears as well as at the tips of the teeth. Because clearances cannot be held lower in a small pump than in a large one, internal leakage is larger fraction of the output flow in a small pump. For this reason, small-size gear pumps are less efficient than large pumps and are rated at lower pressures. The maximum efficiency obtained from a large gear pump is about 80 percent. Small-pump efficiencies are below this figure. Leakage is also largely dependent upon operating pressure and the viscosity of the hydraulic fluid. The

pump manufacturer's recommendations should be followed when choosing a hydraulic fluid. If a particular hydraulic fluid has been specified, a pump designed to operate efficiently with that fluid should be used.

In some gear pumps, the pressure on opposite sides of the gears is equalized by means of passages cut into the shaft of the driven gear. This prevents the gears from being forced against one side of the housing by unbalanced pressure, thereby decreasing wear and increasing efficiency. Some gear pumps are equipped with "wear plates" which are forced against the sides of the gears by discharge pressure. Wear plates decrease oil leakage, but they also increase mechanical friction within the pump, necessitating more driving power and increasing the wear rate. In some gear-pump designs, discharge pressure is used to lubricate the gear-shaft bearings through small holes in the housing between the discharge port and the bearings.

18-1.4 Vane Pumps

The simplest form of vane pump (Fig. 18-2) consists of a stationary housing (stator) within which is a rotating, slotted cylinder (rotor) fitted with movable, radial vanes. The rotor is mounted eccentrically in the cylindrical interior (bore) of the stator. As the rotor turns, the vanes are forced against the hardened-steel ring (cam) on the outer periphery of the bore by centrifugal force, assisted by discharge pressure. In this way, the vanes move the oil from the inlet port to the discharge port.

Most vane-pump designs specify hardened steel for the slotted rotor and the cam. High-speed steel is specified for the vanes in order to prevent rapid wear caused by heating of the vane tips as they rub against the cam.

During each revolution of the rotor, the eccentricity of the device forces the vanes in and out of their slots. Thus, the volume of oil between any two vanes decreases and increases accordingly. The ports, at the top and bottom of the stator bore, extend along arcs concentric with the circumference of

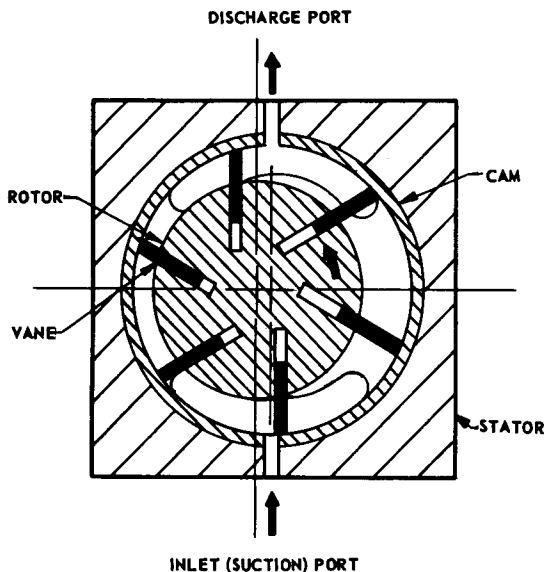


Fig. 18-2 Vane pump.

the bore. The angular distance between the ports is made slightly greater than the angle between two adjacent vanes to prevent intermittent direct connection between the ports. In some vane-pump designs, the stroke can be varied by shifting the stator in relation to the rotor, thereby varying the eccentricity.

High radial loads act on the shaft bearings of the vane pump shown in Fig. 18-2 because the discharge pressure prevails over one half of the rotor. Some vane-pump designs avoid these radial loads by using a double cam in the housing, as well as dual and diametrically opposite inlet and discharge ports. In such a design, each vane passes through two suction and two discharge cycles during each shaft revolution. Since the two rotor areas subjected to high pressure are equal and since the forces acting on the areas oppose each other, no net force acts on the shaft bearings.

The rated discharge of vane pumps ranges from 1 to 60 gpm. Continuous-pressure ratings are usually about 1000 psi; intermittent ratings run to 1500 psi. Some vane pumps

have been built that could deliver 3000 psi. The maximum over-all efficiency of the larger sizes of vane pumps is about 85 percent; smaller vane pumps are somewhat less efficient.

18-1.5 Piston Pumps

Piston pumps are almost always selected for auxiliary service when the required delivery pressure is more than 1500 psi. They may be used either as adjustable-delivery or fixed-delivery pumps. Their maximum deliveries range from 1 to 60 gpm. Rated pressures are as high as 5000 psi. The maximum mechanical efficiency obtained is about 85 percent for pressures below 2000 psi and usually less than that above 2000 psi. There is no significant difference between the characteristics of axial and radial piston pumps. The operation of piston pumps is described in Par. 13-6.

18-1.6 MAINTENANCE

Most hydraulic pumps used on ordnance equipment evolved from industrial designs. Since industrial pumps are subjected to a use more severe than that found in most military systems, they are usually rugged enough for most types of military service, provided they receive an adequate supply of clean oil. The clean-oil requirement is the price that must be paid for continuous, trouble-free operation. Foreign particles in the oil abrade the close-fitting operating surfaces of a pump and ruin them. Excessive heating can also produce this result. Routine maintenance consists almost entirely of periodic inspection of the oil supply and filters. The oil supply should be maintained at an adequate level; the filters should be cleaned regularly and replaced when they no longer perform their function.

18-1.7 LEAKAGE AND DRAINAGE

A properly assembled pump should not leak at any point, except possibly around the drive shaft where it is almost impossible to make an absolutely oil-tight seal. Most pumps are fitted with shaft seals which control oil leakage. These seals are usually not intended

to sustain case pressures of more than 10 psi but, in fact, depend upon the hydraulic fluid in the pump for lubrication of the shaft (see Par. 18-3). Most pumps are also fitted with drain ports to prevent build-up of excessive pressure within the pump case. Drain ports should be vented to the system sump from the top of the pump case. The drain ports should not be vented to the

suction line because the vacuum of the suction line may entrain pump-case air in the oil.

18-1.8 COST

Of the three types of pumps described above, vane pumps are the least expensive and piston pumps are the most expensive. The cost of gear pumps is between the costs of the other two.

18-2 HYDRAULIC AUXILIARIES

18-2.1 HYDRAULIC SYSTEMS INCORPORATING AUXILIARIES⁽¹⁾

Figure 18-3 shows the hydraulic circuit and auxiliaries that are commonly used in a hydraulic transmission. Relief valves are connected across the output of the variable-delivery main pump to limit pressure if the hydraulic motor is overloaded. A fixed-delivery auxiliary pump supplies fluid to the hydraulic stroke control and replenishes any leakage from the main pump and hydraulic motor. Pressure to the hydraulic stroke control is held at pressure P_1 by a pressure regulator, while the pump-replenishing pressure is held at pressure P_2 by a throttling pressure regulator. Replenishing pressure is admitted to the main pump lines through check valves to make sure that it is admitted to the low-pressure line. In practice, the throttling pressure regulator is sometimes omitted and the hydraulic stroke control pressure is also used for replenishing.

The auxiliaries used in a control-valve and hydraulic-ram system are shown in Fig. 18-4. Fluid is supplied by a constant-delivery pump to an unloading valve. An accumulator performs two functions: (1) it stores fluid under pressure when system demand is low; and (2) it supplies fluid to the system when the unloading valve is returning pump delivery to the sump or when the system demand cannot be met by the pump alone.

Pressure-relief valves prevent excessive pressures in the lines to the ram.

18-2.2 CHECK VALVES

Check valves are automatic devices that permit a liquid or gas under pressure to flow through them only in one direction. In the flow direction, the valves open when the applied pressure reaches a small predetermined level and close again when the pressure drops below that level. Pressure in the opposite direction reinforces the force that keeps the valve closed.

18-2.3 Ball Check Valves

Figure 18-5 illustrates the ball check type of valve. When fluid under pressure is applied to port C , ball A is forced upward against compression spring B , and the fluid flows through the valve and out through port D , as the flow arrows indicate. If the pressurized fluid were applied to port D instead of to port C , ball A would be held more firmly in place on its seat and no flow would result. Valve spring B is made with a low spring constant and is installed with only a slight preload to minimize pressure drop through the valve. Other types of check valves operate on the same principle as the ball check valve, but use flat buttons instead of balls, and either lapped or resilient-plastic seating surfaces.

AUXILIARIES ASSOCIATED WITH SERVOMECHANISMS

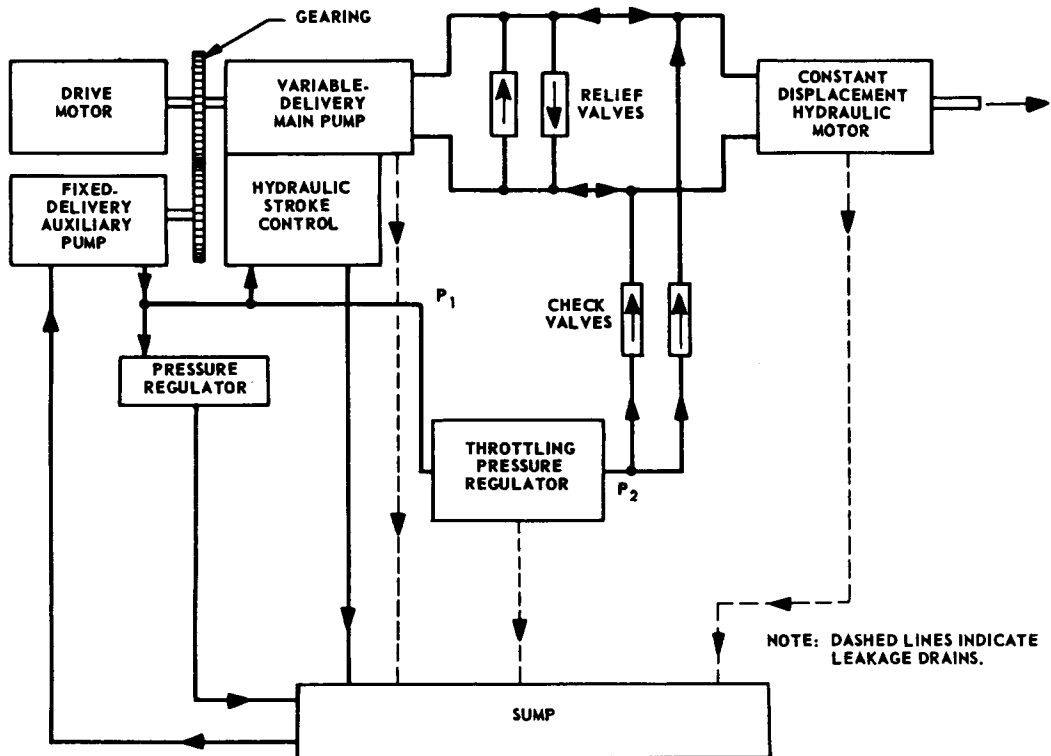


Fig. 18-3 Typical auxiliaries used with hydraulic transmission.

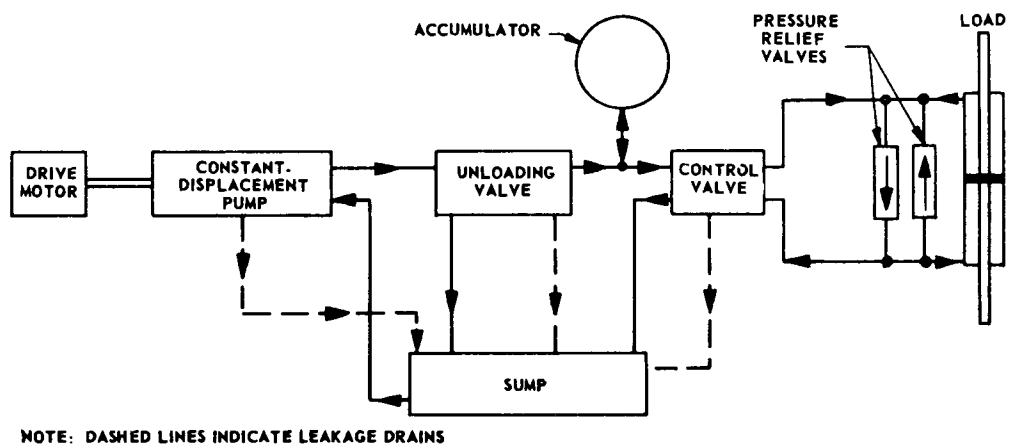


Fig. 18-4 Auxiliaries used in control-valve and hydraulic-ram system.

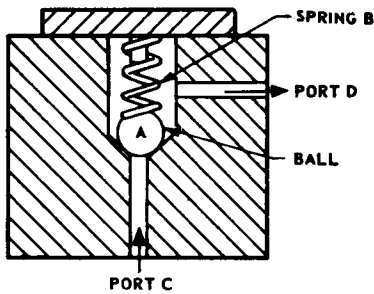


Fig. 18-5 Ball check valve or relief valve.

18-2.4 PRESSURE-RELIEF VALVES

For equipment protection, the pressure in hydraulic lines is relieved, or prevented from exceeding a predetermined value, by pressure-relief valves. Ball relief valves (Fig. 18-5) are frequently used for this purpose. The ball spring is preloaded to the point where the valve will open when the pressure reaches the predetermined value. The valve opens when

$$P = \frac{F}{A} \quad (18-1)$$

where

P = pressure at port C above the pressure at port D, in psi

A = cross-sectional area of the ball at its seating surface, in in.^2

F = spring preload, in lb

As the flow increases, the pressure at port C (Fig. 18-5) rises somewhat, because the spring must be compressed to accommodate the higher flow and because the high velocity of the oil between the ball and its seat reduces the pressure applied to a small area of the ball surface.

For prevention of damage, it is desirable to limit the pressure in the two main lines of a hydraulic system and to permit excess oil

to flow from the high-pressure line to the low-pressure line (see Fig. 18-4). Two ball relief valves are often used for this purpose and are connected between the main lines so that they permit flow in opposite directions. The main line carrying the higher pressure (beyond the predetermined level) discharges the excess oil to the other main line through the ball check valve connected for that flow direction.

A double-acting relief valve (Fig. 18-6) can be used in place of the two ball check valves discussed above. In Fig. 18-6, ports A and B are connected to the main lines to be relieved and the drain is connected to the system sump. Pressure P_1 is applied over a piston area of $\pi(a/2)^2$, and pressure P_2 over a piston area of $(\pi/4)(b^2 - a^2)$. If pressure P_1 is higher than pressure P_2 and also high enough to force the piston downward against the compression spring, oil will flow from port A to port B when the passage between the ports is opened. The flow is in the opposite direction (from port B to port A) if pressure P_2 is higher than pressure P_1 and also high enough to force the piston down far enough to open the passage between the ports.

18-2.5 PRESSURE-REGULATING VALVES

In some small hydraulic systems, it is economically advantageous to use a single fixed-delivery pump that has a delivery rate slightly higher than the maximum system demand. The pressure in the system is controlled by a pressure-regulating valve similar to that in Fig. 18-7. When the hydraulic system uses the entire pump delivery, keeping pump pressure low, the valve is held in the closed position by the compression spring. When the system uses only part of the delivery, the pump pressure rises and the piston is forced downward against spring tension until the port connected to the sump is uncovered. In this way, the unused portion of the delivery is returned to the system sump. If none of the delivery is used by the system, the entire delivery returns to the sump through the pressure-regulating valve.

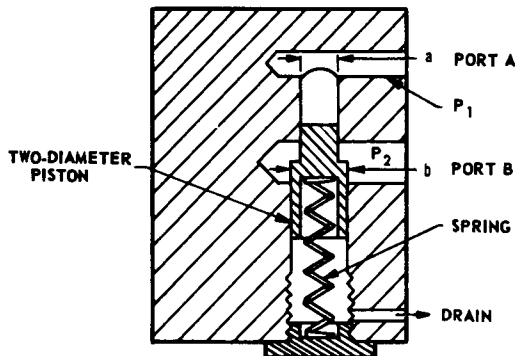


Fig. 18-6 Double-acting relief valve.

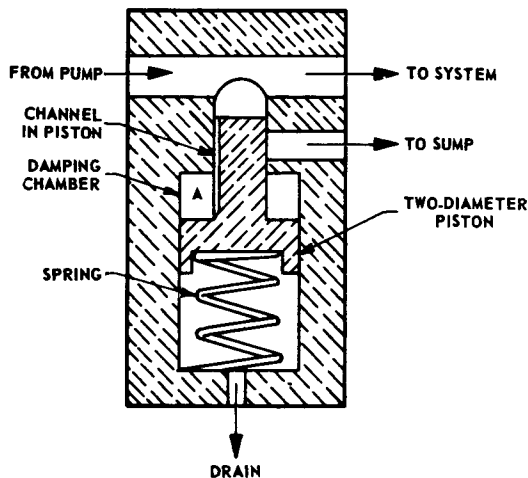


Fig. 18-7 Pressure-regulating valve.

One problem presented to the designer of the pressure-regulating valve shown in Fig. 18-7 is the noise produced by the mechanical oscillation of the piston about its regulating position. Oscillations of the piston are reduced by means of damping chamber A formed by using a valve piston with two diameters. The damping chamber is connected to the pressure side of the piston by

a small channel cut along the length of the smaller part of the piston. When the piston is forced downward by oil pressure, a small amount of this pressure is used to force oil through the small channel into the damping chamber, thereby slightly reducing the pressure on the small diameter of the piston. At the same time, the damping chamber is filled with oil. After the excess delivery has been discharged to the sump, the normal tendency of the compression spring is to rapidly force the piston upward. However, the oil trapped in the damping chamber opposes this action by compelling the spring-driven piston to force the trapped oil through the small channel in the piston. In this way, rapid motion of the piston is damped out. The channel is small enough to stop oscillations of the piston, yet large enough to permit adequate regulator response speed. Actual channel dimensions are usually determined by experiment.

Ball check valves (Fig. 18-5) are sometimes used as pressure regulators, but are almost invariably noisy. However, some commercial regulators use ball check valves as pilot valves to actuate the main regulator. This combination of valves is usually quiet, but the speed of response may not be rapid enough for some applications.

One important disadvantage of the pressure-regulating valve of Fig. 18-7, as well as the ball check valve used as a pressure regulator, is the amount of heat generated in the pumping of oil not used by the system. The work lost in generating this heat reduces the efficiency of a system using one of these pressure-regulating valves. In addition, the heat must be dissipated in some manner, either through water-cooled heat exchangers which occupy relatively little space or through air-cooling of the sump. The air-cooling method must be used in most ordinance equipment because water is unavailable.

A throttling pressure regulator (Fig. 18-8) drops pressure from a regulated high pressure to a regulated lower pressure. When the lower pressure is too low, the compression

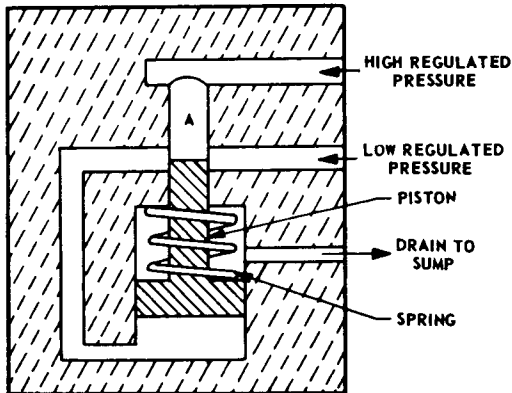


Fig. 18-8 Throttling pressure regulator.

spring opens the valve by pushing the piston downward until passage A between the high-pressure and low-pressure ports of the valve is open. This permits the high-pressure oil to flow to the low-pressure side of the system. When the pressure in the low-pressure side reaches its normal value, the piston is forced upward, closing the passage. This type of regulator permits no more oil to flow from the high-pressure side than that required by the low-pressure side (the load). For this reason, it is not suitable for use as the sole regulator of the output of a fixed-delivery pump. When supplied with a regulated pressure of approximately 500 psi, throttling pressure regulators of this type can be used to regulate the replenishing pressure (approximately 50 psi) of a hydraulic transmission.

18-2.6 ACCUMULATORS⁽²⁾

Accumulators are used to store hydraulic fluid under pressure. They therefore contain stored energy that is available on demand. Accumulators are used in some guided missiles as the sole source of energy used by the control system of the missile during flight. In other applications, accumulators store energy when the system demand is low

and make it available for use when the system demand suddenly increases, thereby decreasing the average power demand of the system.

18-2.7 Gravity Accumulators

The simplest form of accumulator (Fig. 18-9) consists of: (1) a vertical cylinder with a port at the bottom; (2) a single-ended piston inside the cylinder, with the piston rod pointed upward; and (3) a weight secured to the piston rod. Oil under pressure is forced into the cylinder port, pushing the weighted piston upward to its limit of travel in the cylinder. Assuming there are no friction losses, the stored energy thus made available is the product of the weight (in lb, oz, etc.) and the downward distance the weighted piston travels to its lower limit when discharging the oil to the system. The gravity accumulator is suitable only for fixed installations and normally only for low pressures. However, It does have the advantage that its discharge pressure is constant for the entire distance of piston travel. Another

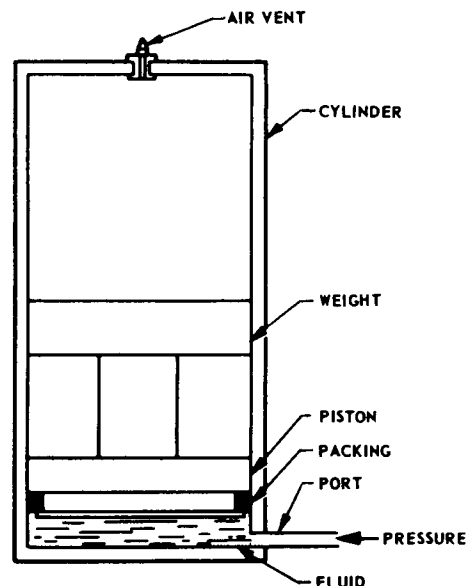


Fig. 18-9 Gravity accumulator.

version of the gravity accumulator uses compression springs instead of a weight, but this type is usually large and high pressures are difficult to obtain.

18-2.8 Hydropneumatic Accumulators

Gas-filled, or hydropneumatic, accumulators are often used in ordnance equipment because of their small size and high pressure and because of the large quantity of available energy that they provide.

A piston-type hydropneumatic accumulator (Fig. 18-10A) usually consists of a cylinder within which the gas and fluid are confined, and uses either a single-piece piston or a pair of opposed pistons operating in cylinder bores of different diameters. A drain is usually provided, as shown in Fig. 18-10A,

to avoid possible seepage of hydraulic fluid into the pneumatic system, or vice versa. The gas is first brought up to specified pressure and the gas shutoff valve is closed. Then the fluid is brought up to pressure, thereby forcing the piston to compress the gas still further. The relation between the two pressures, assuming no friction at the piston seals, is

$$P_g A_g = P_h A_h \quad (18-2)$$

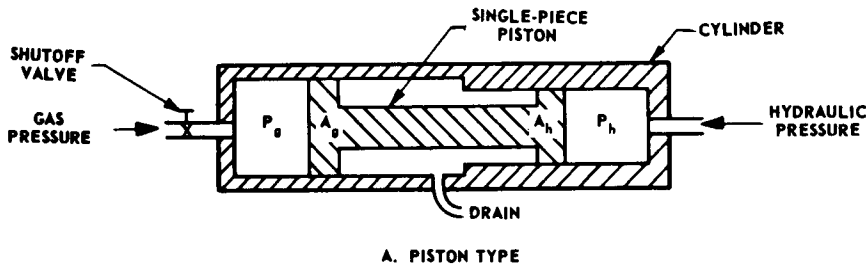
where

P_h = hydraulic pressure, in psi

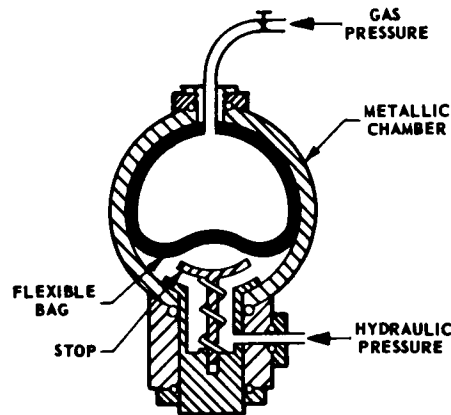
P_g = gas pressure, in psi

A_h = area of hydraulic piston, in in.²

A_g = area of gas piston, in in.²



A. PISTON TYPE



B. BAG TYPE

Fig. 18-10 Hydropneumatic accumulators.

Assuming that the gas expansion is isothermal, the energy available when the accumulator is discharged can be expressed by

$$W = P_o V_o \ln \left(\frac{P_1}{P_2} \right) \quad (18-3)$$

where

W = energy released, in lb-in.

P_o = pressure in gas chamber when fluid chamber is empty, in psi

V_o = maximum volume of gas chamber, in in.³

P_1 = initial gas pressure, in psi

P_2 = final gas pressure, in psi

\ln = natural logarithm (base e)

Piston-type hydropneumatic accumulators have been made with capacities up to approximately 230 in.³, for a nominal system pressure of 3000 psi, but there is no reason why accumulators with larger capacities could not be made. For system pressures much above 3000 psi, however, providing adequate piston seals in this type of accumulator is a difficult problem.

The bag-type hydropneumatic accumulator (Fig. 18-10B) consists of an outer metallic chamber and an inner flexible synthetic-rubber bag which separates the gas and the fluid. Customarily, the chamber is spherical or cylindrical with hemispherical ends. When an initial charge of gas is placed in the bag, a spring-actuated piston-type stop covers the hydraulic port to prevent extrusion of the bag into the port. Then hydraulic fluid is forced into the space between the bag and the interior wall of the chamber, thus producing higher compression of the gas within the bag, until the gas and hydraulic pressures are about equal. The stored energy thus obtained is the same as in the piston-type accumulator. Commercial bag-type accumulators are available with total capacities ranging from 2 to 2300 in.³; maximum pressure is 6000 psi.

Low-pressure hydropneumatic accumulators have been made in which the gas and fluid are in direct contact (no bag or piston is used). This type of accumulator is not generally recommended because the gas dissolves in the fluid at high pressure and comes out of solution when the pressure is removed, producing gas bubbles that may cause serious difficulties in the hydraulic system.

18-2.9 UNLOADING VALVES

Unloading valves (Fig. 18-11) are devices that: (1) permit delivery of pumped oil to the load when the load pressure is below a particular value; (2) permit return of pumped oil to the sump at a low pressure when the load pressure is above a preset value; and (3) prevent return of the system fluid to the sump when the pump is unloaded. Unloading valves are generally used with an accumulator connected to the load to provide a source of available power while the pumped oil is being returned to the sump. Unloading valves actually comprise a number of components within a common housing. In the type shown in Fig. 18-11, port P is connected to the pump, port S to the load, and port R to the sump. Port D is also connected to the sump, but through a separate line. When the system pressure is low, pump pressure opens check valve C which permits oil to flow to the system. Pilot valve V is held against its smaller upper seat by compression spring A , thereby connecting piston chamber L to port D . Bypass valve M is held against its seat by compression spring B and, consequently, full pump delivery flows to the system.

When the system pressure rises above a predetermined value, it forces pilot valve V downward against the force of compression spring A (Fig. 18-11). Since the lower sphere and seat diameters of pilot valve V are larger than those of the upper sphere and seat, entrance of pressure into the upper valve chamber forces V to close on its larger seat. Oil from the system pressure line then flows into piston chamber L , forcing bypass valve M downward against the force of spring B .

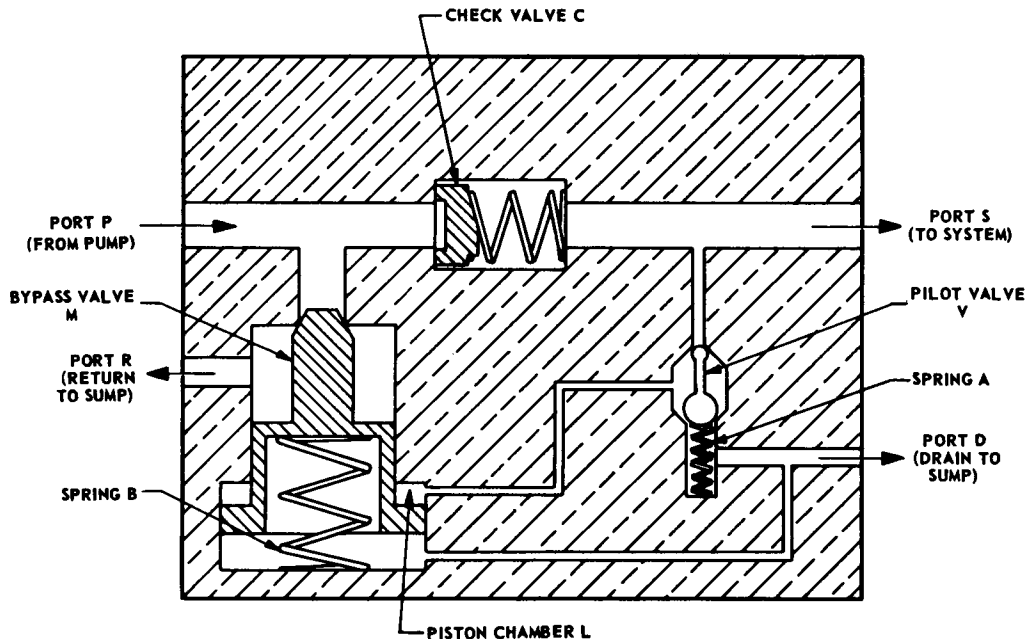


Fig. 18-11 Unloading valve.

Pumped oil is now free to flow back to the sump from port *P*, past bypass valve *M*, and out through port *R*. Check valve *C* closes (as shown in Fig. 18-11) because the pumped oil is bypassed to the sump, making pump pressure lower than the system pressure.

When the system pressure drops low enough to permit spring *A* to lift pilot valve *V* from its larger seat, *V* closes again on its smaller seat because of the valve proportions; piston chamber *L* is connected to port *D* through *V*; spring *B* forces bypass valve *M*

to close the passage between *P* and *R*; and pump pressure opens check valve *C*, permitting delivery to the system.

In some unloading valves of the type shown in Fig. 18-11, the passage to chamber *L* is restricted to prevent flutter of bypass valve *M*. Most unloading valves require extremely clean oil because even the slightest amount of foreign matter in the oil could prevent proper operation of pilot valve *V*, which moves only a few thousandths of an inch between its upper and lower seats.

18-3 ROTARY JOINTS

18-3.1 DYNAMIC SEALS

Dynamic seals must be used when a rotating or translating shaft must pass through the wall of a chamber in which fluid must be confined. To obtain a reasonable life expectancy from the seal, it must be lubricated by the confined fluid. For this reason, successful seals always leak enough fluid to keep the shaft moist at the sealing surface, but this leakage should be extremely small.

18-3.2 Glands

One of the oldest and simplest forms of dynamic seal is the stuffing box or packing gland shown in Fig. 18-12. A soft, compressible material (for hydraulic fluid this is usually lead wool or asbestos) is packed into the hollow recess of the gland and compressed by the screw. These seals usually have relatively high friction and leakage, and their use is confined to applications involving low-speed shafts with rotary or translatory motion. When properly installed, glands are effective for low-pressure applications. In most cases, glands are considerably larger than some other types of seals.

18-3.3 O-Rings^(3, 6, 7)

O-rings are employed primarily to seal translating shafts. They are not recommended for high-speed continuously rotating shafts, although they have been used for sealing oscillating shafts that rotate a maximum of one revolution or less. As some compression of the O-ring is required to produce a seal, some friction is always present, even with zero pressure differential across the ring. For this reason, O-rings should not be used when friction forces must be minimized. Compared with O-rings, U-cups and V-ring seals produce somewhat lower friction forces at low pressure differentials. Two applications of O-rings are shown in Fig. 18-13. The dimensions of standard O-rings are given in Air Force-Navy Aeronautical Standard AN6227. O-rings are usually fitted

into rectangular grooves. For applications similar to the piston seal of Fig. 18-13A, O-rings can be used with leather backing rings to increase the permissible pressure differential across the piston from 1500 psi without the backing ring to 3000 psi with the backing ring. When O-rings are being installed, they must be protected from scratches or cuts because even small surface defects produce high oil leakage. The surface against which the O-ring slides, as well as the bottom of the groove in which it is installed, must be well-polished.

18-3.4 U-Cup and V-Ring Packings

These types of packings are commonly used to seal translating surfaces. Applications and characteristics of U-cup and V-ring packings are given in the Mechanical Engineer's Handbook.⁽⁴⁾

18-3.5 Shaft Seals⁽⁵⁾

Shaft seals are used to seal continuously rotating shafts when the pressure is below 35 psi. Three types of construction are in general use (Fig. 18-14). In the construction

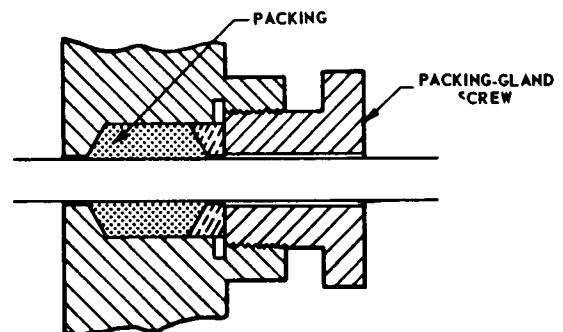


Fig. 18-12 Packing gland.

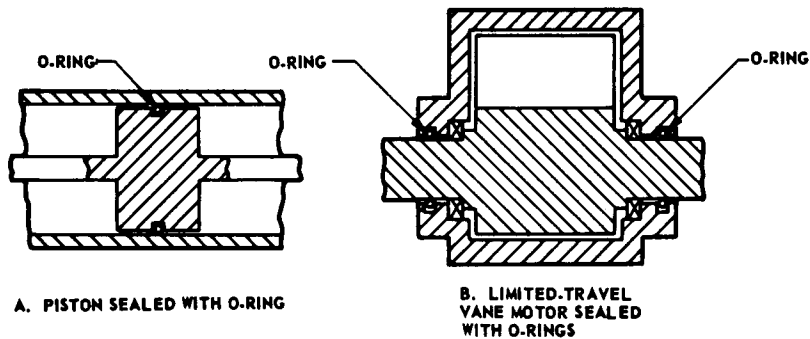


Fig. 18-13 Applications of O-rings.

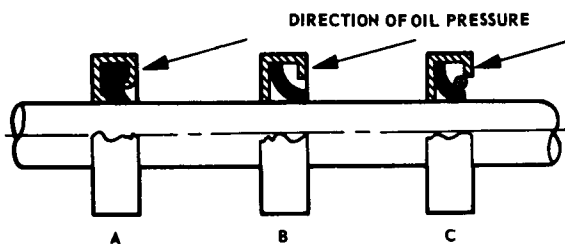


Fig. 18-14 Shaft seals.

illustrated in Fig. 18-14A, no back-up spring is used. The elasticity of the sealing material is utilized to keep the sealing material in contact with the shaft. In Fig. 18-14B, leaf-spring fingers are used to reinforce the sealing material elasticity. In Fig. 18-14C, a helical spring formed into a circle serves to keep the sealing lip in contact with the shaft. There is little difference in sealing efficiency between these three types. The sealing-lip material is usually leather or synthetic rubber.

Installation of shaft seals requires great care in order to avoid even the slightest damage to the sealing lips. An arbor or collar, with one end rounded or tapered, is

sometimes used to open the seal to the shaft diameter and to start it over a smooth portion of the shaft. Before shaft seals with synthetic-rubber sealing lips are installed, both the shaft and the installing tool must be oiled. Shaft seals are installed by pressing the external metal shell of the seal into a recess in the housing, with the seal oriented so that pressure of the fluid being sealed forces the sealing lip against the shaft (unless the pressure is very low).

18-3.6 Face Seals

High-speed shafts are sometimes sealed with face seals (Fig. 18-15) when the oil pressure ranges as high as 2000 psi. In this type of seal, two materials that have a low coefficient of friction are used to establish the dynamic seal and the dynamic sealing surfaces are lapped to obtain a very smooth finish. The static seal between the shaft and the rotating face of the dynamic seal must be flexible enough to accommodate a small amount of runout between the surfaces of the dynamic seal. Many arrangements are used to drive the rotating face and to provide the static seal. The dynamic sealing surfaces must be lubricated at all times; otherwise, the face seal will be damaged quickly by the heat resulting from dry friction. Descriptions of seals used to prevent low-pressure

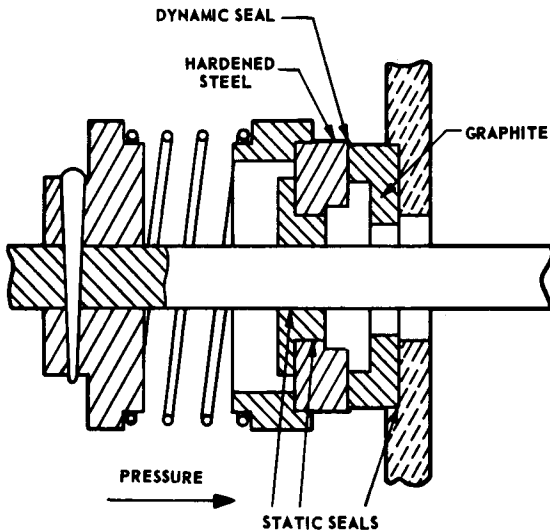


Fig. 18-15 Face seal.

gas from leaking past a small shaft are given in references 8 and 9.

18-3.7 High-Pressure Seals

For high-pressure seals with low friction, an arrangement similar to that in Fig. 18-16 is sometimes used. This seal is not recommended for high-speed shafts or in applications requiring extremely low leakage from the pressure vessel. The seal will reduce leakage from the end of the shaft to a very small amount. Basically, the seal comprises a series of grooves around the shaft approximately 1/16-inch wide and 1/32-inch deep. The effect of the grooves is to equalize the pressure around the entire shaft circumference and to keep the shaft centered in the housing. Oil leaking past the grooves is carried away by a drain line to the sump. Oil flow from the end of the shaft is prevented by a commercial shaft seal. The shaft and housing should be made of materials having low coefficients of friction, preferably hardened

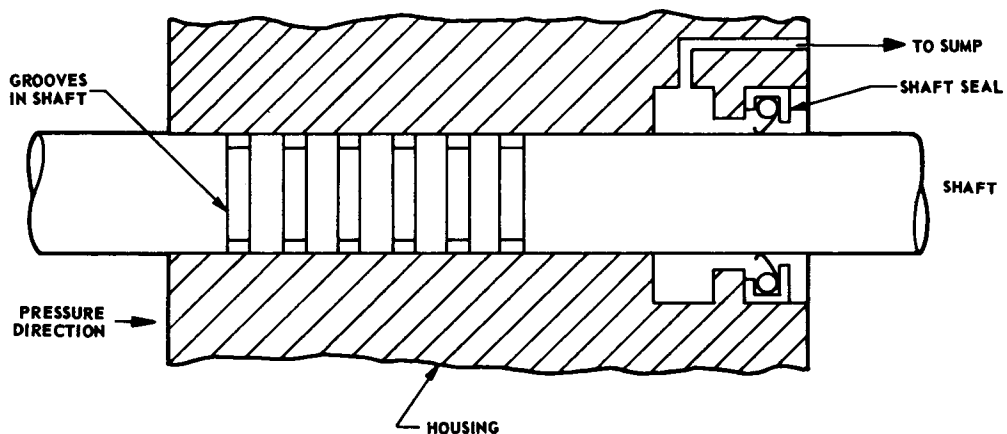


Fig. 18-16 High-pressure seal.

steel which also prevents foreign particles from being embedded in the material. Clearance between the shaft and its passage in the housing should be low, and the edges of the grooves at the shaft surface should be sharp and clean.

18-3.8 Friction

All dynamic seals produce some friction and should not be used on shafts where friction must be kept to an absolute minimum. For example, the shafts of torque motors or free synchro receivers should not be sealed. These devices, provided they have no slip rings, can usually be submerged in oil with no ill effects; in fact, submerging them in oil appreciably decreases their temperature rise.

Little quantitative information about seal friction is available, probably because of the wide variations that occur between seals that should produce the same frictional torque or force. Nevertheless, the following general comments can be made:

(a) Synthetic-rubber packings have a tendency to break through the lubricating-oil film while a shaft is at rest and, consequently, may exhibit an initial starting

friction that is considerably higher than the subsequent running friction. This tendency is much less pronounced with leather packings, possibly because leather can absorb oil and thus become partially self-lubricating.

(b) A smooth surface on the moving member of the seal produces less friction than a rough surface.

(c) Synthetic-rubber seals cannot be used with fluids that produce excessive swelling of the rubber because high friction and rapid wear will result.

(d) For O-rings:⁽⁷⁾

(1) The use of leather backing rings increases seal friction on the shaft.

(2) The squeeze of the O-ring should not be excessive.

(3) The clearance between the shaft and the bore in which the O-ring moves should be held as low as practicable on the low-pressure side to prevent extrusion of the ring into the clearance.

(4) O-rings with a durometer hardness reading of approximately 65 have somewhat lower friction than O-rings with a durometer hardness reading of approximately 90, provided there is no extrusion.

18-4 LIMIT STOPS AND POSITIVE STOPS

18-4.1 PURPOSE

Stops are often desirable or even necessary in servomechanisms to limit output motion or travel. For example, azimuth drives of anti-aircraft guns might well be limited to avoid the possibility of firing in a direction that would cause shells to fall on friendly troops. The elevation motions of tank and artillery guns are always limited, and stops are necessary to prevent equipment damage that might occur if it were possible to drive the guns into the travel limits at full speed and torque.

18-4.2 CHARACTERISTICS

Limiting stops may be either positive stops, which halt the output motion by overpowering the servomotor, or limit stops. Limit stops disconnect the normal servomechanism error signal and replace it with a signal originated by the output motion, which causes the output to stop. Because positive stops are easy to make, and operate directly on the output, some devices are equipped with both positive and limit stops. Positive stops are often necessary to avoid damage during manual operation of a device

AUXILIARIES ASSOCIATED WITH SERVOMECHANISMS

servomotor actuates the arm of a spring-centered potentiometer, near either end of the prescribed output travel. The output voltage of the potentiometer network is zero when the potentiometer arm is centered and increases when the arm is moved in either direction from center. The phase or polarity of the output voltage is governed by the direction of arm movement.

The maximum output voltage of the potentiometer network (Fig. 18-18) is higher than the maximum output voltage of the saturating error amplifier. The potentiometer and the saturating error amplifier are both connected to the output-amplifier input through a summing network, the entire arrangement being set up in such a way that the potentiometer-network voltage always subtracts from the error voltage. When servomotor rotation nears either limit, the cam moves the potentiometer arm, thereby increasing the potentiometer-network output voltage. Thus, the voltage input to the output amplifier is reduced to zero and the servomotor is stopped. This limit-stop system is a position servomechanism that uses the output of the saturating error amplifier as its command signal. Therefore, the exact position at which the servo output stops depends

upon the magnitude of the saturating-amplifier output voltage. This system operates with little difficulty if d-c voltages are being summed. However, some difficulty can be expected if a-c voltages are used, because of the phasing of the two signals and because the waveforms of the mixed signals differ (the output of the saturating error amplifier is clipped, while the output of the potentiometer network is essentially sinusoidal).

A mechanical limit-stop system, a form of position servo, is shown in Fig. 18-19. Error motion at a reasonably high force level is transmitted through both the spring-loaded force limiter and the box around the spring-centered limit-stop arm to the control valve. The valve controls the direction of rotation as well as the starting and stopping of the hydraulic motor, thereby controlling the position of the cam that actuates the limit-stop arm. When the cam is well clear of the limits, error motion is transmitted without interference to the control valve, which opens one valve port and causes rotation of the motor. If the cam eventually turns the limit-stop arm to one side, the boss on the arm shifts its enclosing box, forcing the control-valve piston back to neutral. This stops the hydraulic motor. The control valve

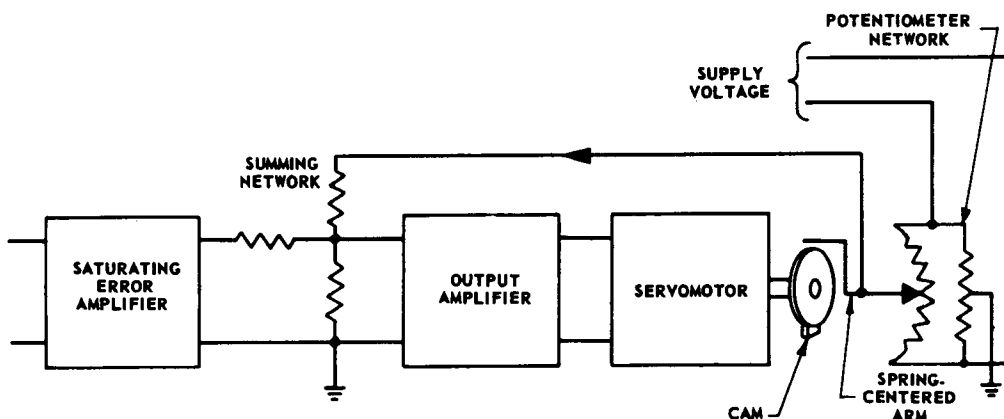


Fig. 18-18 Electrical limit-stop system.

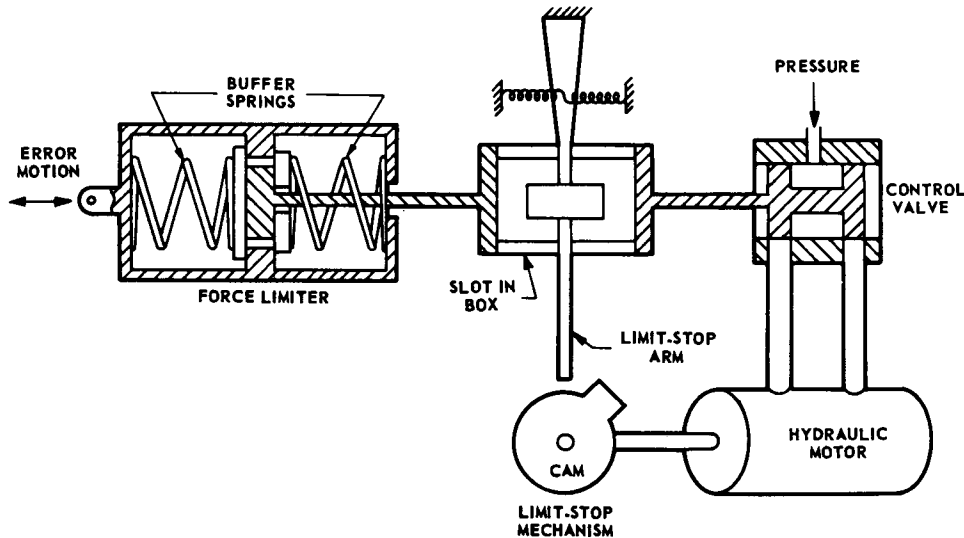


Fig. 18-19 Mechanical limit-stop system.

can be moved far enough to stop the motor without damaging the error-input mechanism because of the spring buffers provided in the force limiter. If error motion is obtained from a source that is not damaged when it is forcibly reversed against its full torque (as would be the case if a synchro or torque motor were used), the force limiter can be omitted.

When the limit-stop arm is actuating the control valve, the motor, cam, and valve comprise a closed loop which can be considered as a position servomechanism. As a result, the mechanical limit-stop system must be designed to suit the requirements of the servomechanism with which it is to operate in order to obtain stability (nonoscillating operation) of the device at the limits.

In general, limit stops should be designed to rely on a minimum number of other devices for their operation. Electronic amplifiers or other devices that can be expected to fail occasionally should not be placed in the limit-stop closed loop.

18-4.4 POSITIVE STOPS

The simplest form of positive stop consists of a lug, attached to the moving member, that strikes a fixed member at the end of the desired travel (Fig. 18-20). If the output motion exceeds one revolution, lugged washers can be stacked on the output shaft so that their lugs overlap, as shown in Fig. 18-21. The lug of the washer at one end of the stack is engaged by a lug on the shaft. As the shaft rotates, the lug pick-up action is cumulative until the lug on the washer at the other end of the stack is engaged by a fixed block or pin at the completion of the required number of revolutions. Lugged washers are often used on instrument servomechanisms.

18-4.5 Buffers

Springs are often incorporated in positive stops to produce buffer action that reduces the acceleration produced when the stops are engaged. Often, the addition of a piece of thick synthetic rubber to each side of the

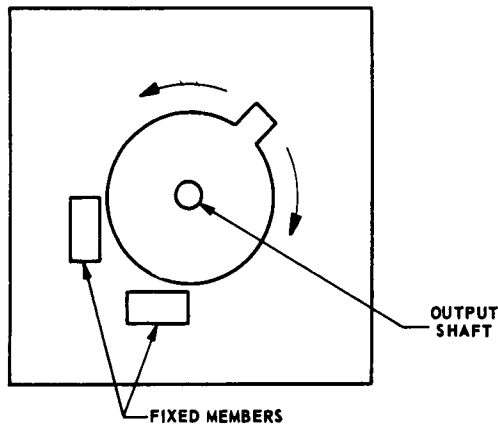


Fig. 18-20 Positive stop.

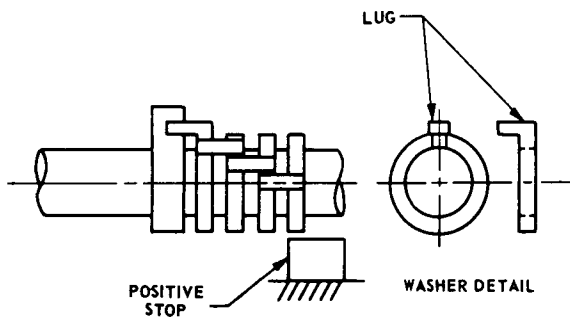


Fig. 18-21 Stop with lugged washers.

lug arrangement of Fig. 18-20 produces sufficient shock-absorbing action to prevent damage. In other applications, it is necessary to resort to a combination of spring and dash-pot type damping, similar to the system of Fig. 18-22, to obtain the desired deceleration without damage.

Hydraulic cylinders can be equipped with buffer stops at each end of piston travel, as shown in Fig. 18-23. In this system, oil is trapped in the recessed end of the piston by

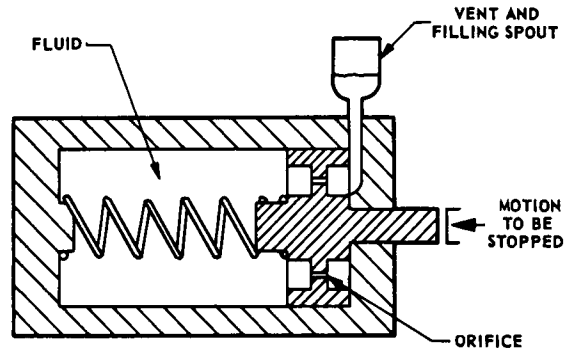


Fig. 18-22 Spring and buffer stop.

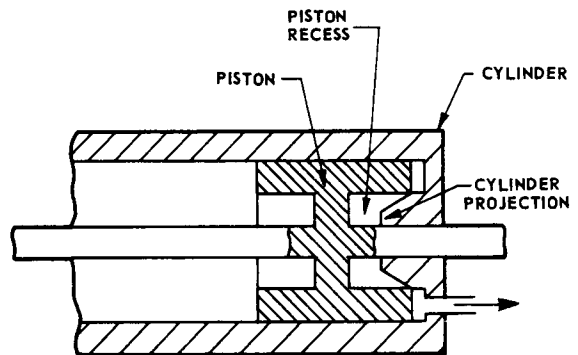


Fig. 18-23 Buffer stop in hydraulic cylinder.

the cylinder projection as the piston nears the end of the cylinder. The trapped oil can escape from the piston recess only through the restricted space between the wall of the piston recess and the surface of the cylinder projection, thereby producing viscous damping of piston movement. Either the cylinder projection or the piston recess, or both, can be shaped to produce any desired retardation rate. Check valves are sometimes used to permit unrestricted flow of oil from the supply line into the damping chamber but to force oil flowing from the chamber to the line to pass through the restricted passage. This arrangement reduces the time required to move the piston away from the end of the cylinder.

BIBLIOGRAPHY

- 1 W. Ernst, *Oil Hydraulic Power and its Industrial Applications*, McGraw-Hill Book Company, Inc., New York, N. Y., 1949.
- 2 E. M. Greer, "Applying Hydraulic Accumulators", *Machine Design*, p. #145, April, 1947, and p. #107, May, 1947.
- 3 Air Force-Navy Aeronautical Standard AN6227, Gasket, O-Ring; AN6246 Ring-Hydraulic Packing Back-up.
- 4 Kent, *Mechanical Engineer's Handbook*, Design and Production Volume (12th edition), John Wiley & Sons, Inc., New York, N. Y., 1950.
- 5 K. P. Mosslander, "New Designs Broaden Scope of Shaft Seals", *Machine Design*, p. #44, November, 1940.
- 6 D. R. Pearl, "O-Ring Seals", *Machine Design*, p. #97, May, 1947.
- 7 C. M. White and D. F. Denny, "The Sealing Mechanism of Flexible Packings", *Ministry of Supply Scientific and Technical Memorandum No. 3/47*, His Majesty's Stationery Shop, London, England, 1948.
- 8 "Development of Rotary and Translatory Seals", Syracuse University Research Institute, Frankford Arsenal Contract DA30-115-ORD-363.
- 9 "High Speed, Rotary Seal, Mechanical Drive", *Final Report No. 26*, Krautter-Weber Tool Co., Newark, N. J., Signal Corps Procurement Agency, Fort Monmouth, N. J., Contract W-36-039-SC-44520.

CHAPTER 19

CONSTRUCTIONAL TECHNIQUES*

19-1 BASIC CONSIDERATIONS

In the design of any servomechanism or control system, a great many factors besides the response characteristics must be considered. Many systems that have adequate response characteristics are completely unsatisfactory because too many adjustments are required, the adjustments must be made too often, maintenance is difficult, or the equipment cannot withstand environmental

conditions. In ordnance equipments, it is particularly important that much attention be paid to these details, not only because the equipment must be made reliable and simple to operate under conditions normally imposed upon equipment and operator, but also to minimize the time required to train operators and maintenance organizations.

19-2 COMPONENT LAYOUT

19-2.1 PHYSICAL ARRANGEMENT

Several physical aspects of component layout should be considered carefully during the design of a servomechanism.^(1,2) Controls that must be used by the operator for successful operation of the device should be accessible at all times. For example, an amplifier balance control that must be adjusted each time the amplifier is turned on, or readjusted as the amplifier warms up, should be easily available to the operator. However, the control should not be placed where it can be moved accidentally. Switches should be placed for convenient use and the switch position (on, off, or other) should be marked. If several switches are included on a panel, the function of the switch should also be marked.

Controls requiring readjustment during maintenance should be inaccessible to operating personnel. However, the controls should be easily accessible to maintenance personnel, perhaps requiring only the removal of a simple cover or cap. It should not be necessary to remove an entire unit to readjust the gain control of the main loop of a servomechanism incorporating an electronic amplifier. Also, this control should not be accessible to the operator so that he can misadjust the system gain to an extremely oscillatory or badly overdamped condition. Access for routine maintenance should require only the removal of a few covers, not the removal of an entire assembly. Also, parts with finite useful lives should be easily replaced. The replacement of electron tubes, for example, should be a simple task. Filters for hydraulic fluid should be arranged so that they can be removed for cleaning or

*By J. O. Silvey

replacement without disturbing anything but the filter element; they should be connected so that only a negligible quantity of fluid is lost when a filter is removed. Controls and adjustments necessary for the initial alignment of the device, but not requiring subsequent readjustment, should be arranged to prevent unintentional readjustment. Furthermore, these controls should be placed where only authorized ordnance maintenance personnel are permitted to make adjustments.

Dials and instruments that must be read quickly and accurately should be placed where they can be seen easily and should be marked distinctively. Considerable psychological testing to determine the most readable instrumentation layouts has been done by the military⁽²⁾. Although the problems encountered in ordnance equipment are not identical with those for which the experiments were made, the general results of the tests are applicable.

Components should be arranged within an assembly so that it can be disassembled and reassembled with no special knowledge or special tools. It is also very desirable that the number of parts be kept as low as practical and that there be only one way of putting the parts together.

19-2.2 THERMAL CONSIDERATIONS AND HEAT GENERATION(3, 4, 5)

Many of the devices used in servomechanisms develop undesired heat which must be dissipated. Where space is not a limitation and air is free to circulate around the device, heat dissipation to the atmosphere may be accomplished with a temperature rise of only a few degrees above the ambient temperature. However, if the volume and hence the area of the devices is small, the resulting temperature rise may be sufficiently high to damage the equipment. The high ambient temperatures that can be encountered by ordnance devices should be considered when designing control systems.

The sources of heat in various control systems differ so greatly from one another that

it is necessary to evaluate each source separately. Some of the most common heat sources, as well as the representative rates at which they liberate heat, are listed in Table 19-1. If the temperature at which a control system of a particular design is to operate appears to be excessive, the heating rates of the actual components should be obtained under service conditions.

The heat generated in a control system must be adequately dissipated to prevent excessive temperature rises. In general, it is preferable that auxiliary cooling devices such as fans and heat exchangers should not be used. However, in some instances, auxiliary cooling devices may be required (for example, to avoid excessive size of the control system).

The mounting of electronic components so that air can flow freely past the heat sources, particularly in an upward direction, is a well-established method of cooling amplifiers. Attaching a heat source (for example, a control motor) to a large plate or structural member of high heat conductivity helps to decrease the temperature rise of the heat source. In hydraulic systems, the temperature of electrical equipment can often be lowered by submerging the electrical device in the hydraulic fluid. Because the fluid is in motion and has much greater thermal capacity than air, the temperature rise of the submerged electrical device will be much less than it would be in air. Normally, hydraulic fluid can be kept adequately cool by using a sufficiently large sump which, in turn, is air cooled.

In most ordnance control equipment, the generated heat is transmitted to the surroundings by radiation or by convection of the air. Unfortunately, the laws of convection and radiation are so complicated that it is not usually practical to establish a single simple relation that will quickly give the heat dissipated by a hot body.

CONSTRUCTIONAL TECHNIQUES

TABLE 19-1 HEAT LIBERATED BY CONTROL DEVICES

| Device | | Rate of Heat Liberation | Remarks |
|---------------------------------------|-----------------------|--------------------------|---|
| Vacuum-tube voltage-amplifier triode | | 4 watts | |
| Vacuum-tube voltage-amplifier pentode | | 3 watts | |
| Vacuum-tube power pentode (Type 6L6) | | 25 watts | |
| Constant-speed 60-cps induction motor | no load | 9% output rating | |
| | full load | 18% output rating | |
| Armature-controlled d-c motor | field only | 6% power rating | |
| | no load | 15% power rating | |
| | full load | 25% power rating | |
| 2-phase control motor | no load | 300% of max output power | Heating with only reference field excited is approximately equal to maximum power output. |
| | stalled rated voltage | 400% of max output power | |
| Auxiliary pump | | 20% of input power | Minimum heat equals 15% of input power. Actual loss depends upon duty cycle and pressure. |
| Hydraulic motor | | 15% of input, minimum | |
| Hydraulic amplifier | | 50% of input power | Minimum heat usually about 30% of input. |
| Variable-delivery pump | | 15% of input, minimum | |

19-2.3 Radiation

Heat radiated by a hot body to a large, cool, black body that completely surrounds the hot body can be determined by

$$Q_1 = 36.9 \times 10^{-12} e (T_1^4 - T_2^4) \quad (19-1)$$

where

Q_1 = radiated heat, in watts/in.²

e = emissivity of hot body (see Table 19-2)

T_1 = absolute temperature of hot body, in °K
(Temperature in degrees Kelvin = 273 + temperature in degrees Centigrade.)

T_2 = absolute temperature of cool surrounding body, in °K

The heat radiated by a hot body can also be determined by

$$Q_2 = 17.2 \times 10^{-10} e (T_a^4 - T_b^4) \quad (19-2)$$

where

Q_2 = radiated heat, in Btu/(hr) (ft²)

e = emissivity of hot body (see Table 19-2)

T_a = absolute temperature of hot body, in °R
(Temperature in degrees Rankine = 460 + temperature in degrees Fahrenheit.)

T_b = absolute temperature of cool surrounding body, in °R

When the radiation loss is to be computed for a device that is enclosed in a chamber, the temperature to be used for T_2 or T_b in Eqs. (19-1) or (19-2) is the temperature of the chamber walls. If the device is outdoors, the temperature to be used for the surroundings is not so easily obtained. The safe approach is to use the temperature of the surrounding outdoor air as the value of T_2 or T_b , although this will, in nearly all cases, yield values for heat dissipation which are lower than those actually encountered. A clear, blue sky reflects little radiant energy. In calculations involving heat dissipation of devices exposed to direct sunlight, it is necessary to dissipate the heat absorbed from solar radiation as well as the heat generated within the device. Neglecting atmospheric absorption and assuming that the solar radiation falls on a black body, the heat resulting from the sunlight is approximately 420 Btu/ft²(hr) if the surface is normal to the solar rays.⁽⁵⁾ Dust, water vapor, ozone, and other substances in the atmosphere reduce the solar heat to a value of approximately 290 Btu/ft²(hr), or lower, at the surface of the earth.

19-2.4 Free Convection by Air

Free convection by air occurs when the air is initially at rest; subsequent air motion results from the expansion produced by the heat extracted from the warm body. In ordinance equipment that is operated outdoors, this type of convection occurs only when there is no wind. Heat loss due to convection

TABLE 19-2 EMISSIVITY OF VARIOUS SURFACES

| Surface (100°F) | Emissivity |
|---------------------------|----------------|
| Polished aluminum | 0.04 |
| Polished copper | 0.03 (approx.) |
| Copper covered with oxide | 0.78 |
| Polished iron | 0.24 |
| Oil paint | 0.94 |
| White paint | 0.90 to 0.95 |
| Black paint | 0.96 |
| Aluminum paint | 0.3 to 0.6* |

*Depending upon amount of aluminum and kind of lacquer.

CONSTRUCTIONAL TECHNIQUES

is dependent upon the attitude of the surface relative to the vertical, the dimensions of the surface, and the temperature difference between the surface and the surrounding air.

For a high vertical surface, the process of free convection by air is somewhat as follows: Cool air is drawn in toward the surface at the bottom. However, because of the viscosity of the air, the layer of air in immediate contact with the surface does not move, forming a boundary layer through which the heat must be conducted. When the air close to the surface becomes heated, it rises slowly, and the flow is laminar or viscous. As the heated air moves upward along the surface, the air becomes increasingly warmer, its cooling ability decreases, and its velocity increases. Eventually, the air flow becomes turbulent, and as cooler air is brought in by the eddies of the turbulence, the cooling ability of the air increases.

Experiments⁽⁴⁾ indicate that heat loss by air convection from a surface may be approximated by the expression

$$H_c = A_c \theta_c^{5/4} \quad (19-3)$$

where

H_c = heat loss, in microcalories/cm²(sec)

θ_c = temperature difference between surface and air, in °C

$$\log (A_c - 50) = 2.22 - \frac{h}{20}$$

h = height of surface, in cm

A more convenient approximate relation for heat loss by air convection is

$$Q_c = A_F H_F \theta_F \quad (19-4)$$

where

Q_c = heat loss, in Btu/hr

A_F = area of vertical surface, in ft²

θ_F = temperature difference between surface and air, in °F

H_F = variable whose value is obtained from Table 19-3

The values of H_F in Table 19-3 are based on values in Table 8B of reference 4 that were obtained from tests made with heated cylinders which were either suspended in air

TABLE 19-3 VALUES OF H_F [H_F in Btu/(hr)(ft²)(°F)]

| Surface Height (in.) | θ_F | | | | | | | |
|-------------------------|------------|------|------|------|-------|-------|-------|-------|
| | 20°F | 40°F | 60°F | 80°F | 100°F | 120°F | 140°F | 180°F |
| 100 | 0.60 | 0.75 | 0.87 | 0.93 | 1.01 | 1.06 | 1.12 | 1.23 |
| 69 | 0.65 | 0.75 | 0.88 | 0.96 | 1.03 | 1.10 | 1.15 | 1.24 |
| 35 | 0.65 | 0.8 | 0.87 | 0.92 | 0.98 | 1.02 | 1.07 | 1.13 |
| 23 | 0.7 | 0.8 | 0.88 | 0.95 | 1.0 | 1.05 | 1.10 | 2.17 |
| 11.3 | 0.75 | 0.9 | 1.00 | 1.06 | 1.12 | 1.17 | 1.23 | 1.32 |
| 5.9 | 1.10 | 1.3 | 1.38 | 1.44 | 1.51 | 1.55 | 1.60 | 1.66 |
| 3.2 | 1.65 | 1.8 | 1.95 | 2.03 | 2.12 | 2.2 | 2.27 | 2.38 |
| 1.8 | 2.0 | 2.4 | 2.55 | 2.70 | 2.86 | 2.98 | 3.10 | 3.28 |

by thin supports or supported by columns no larger in diameter than the heated cylinder. Other tests made with a vertical, flat plate (34 inches square) indicate that a decrease of approximately four percent in convection cooling results when a floor is placed immediately below the vertical plate. The reduction in natural convection cooling that would result from a floor placed immediately below the cylindrical surfaces used for Table 19-3 is not known. A second plate placed parallel to, and at a distance of more than two inches from, a vertical heated surface has no effect on the air convection cooling. Large horizontal heated surfaces facing upward release approximately 30 percent more heat than the vertical cylinders tested for Table 19-3. Also, large horizontal heated surfaces facing downward release approximately 30 percent less heat by convection than the cylinders tested for Table 19-3.

19-2.5 Conduction

Heat transmitted through a material by conduction follows the law

$$Q_T = \frac{AK\theta}{L} \quad (19-5)$$

where

Q_T = heat transmitted, in Btu/hr

A = area of path of heat flow (perpendicular to direction of heat flow), in ft²

K = thermal conductivity of material, in Btu/hr (ft) (F°)
(F° = temperature difference, in °F)

θ = difference in temperature of two sides of material, in °F

L = thickness of material, in ft

The thermal conductivities of some materials are listed in Table 19-4. The conversion factors for the coefficients of heat transfer are given in Table 19-5.

19-2.6 EQUIPMENT AND PERSONNEL SAFETY MEASURES

Ideally, any failure of a servomechanism

TABLE 19-4 THERMAL CONDUCTIVITY OF METALS AND ALLOYS AT 212°F

| Material | Thermal Conductivity [Btu/(hr) (ft) (F°)] |
|-----------------------------|--|
| Aluminum | 119 |
| Brass (70-30) | 60 |
| Copper | 218 |
| Lead | 20 |
| Silver | 238 |
| Steel AISI-1020 | 26 |
| Stainless steel AISI-304 | 9.4 |
| Pine wood — across grain | 0.09 |

should cause an alarm indication and result in complete cessation of the output. Also, the device should be constructed so that personnel injury or equipment damage is impossible. It is difficult, however, to eliminate a servomechanism output for all system failures and to make an alarm system that will indicate failures reliably. The practical requirements that arise in the use of many service equipments make the exercise of common sense and the fulfillment of an adequate training program the only real insurance against injury to personnel. Nevertheless, much can be done to reduce the probability of equipment damage and injury to personnel. Each device should be studied with regard to its specific requirements in order to properly evaluate the safety devices that should be included.

Limit stops and positive stops (Par. 18-4) are often used to prevent damage to equipment. Positive stops are usually accompanied by slip clutches, over-current relays, or hydraulic relief valves so as to avoid damage to the servomotor or associated devices when the servomechanism output strikes the positive stop.

CONSTRUCTIONAL TECHNIQUES

Interlocking relays are often used to prevent operation of a servomechanism when a component has failed; for example, loss of plate voltage to a vacuum-tube amplifier can be used via an interlocking relay to de-energize the motor of an associated amplifier. Solenoid-operated bypass valves and disconnect-linkages are often used to prevent motion of hydraulic or mechanical devices when the servo electric power fails. When it is necessary that the output of the servo go to one limit in case of failure, it is sometimes possible to use relays or spring-loaded solenoids to apply a false error signal that

will cause the output to slew to the proportional limit.

Most power servomechanisms used to drive guns and tank turrets have more than enough torque to crush any portion of the human anatomy that happens to be in the path of the output. It is important that the output be equipped with guards to prevent personnel injuries where moving members are close to fixed members. These guards must not interfere with normal operation of the weapon. Adequate protection against accidental contact with high-voltage electric circuits should also be provided.

TABLE 19-5 CONVERSION FACTORS FOR COEFFICIENTS OF HEAT TRANSFER

| <div style="display: flex; align-items: center; justify-content: center;"> <div style="text-align: right; margin-right: 10px;"> Multiply number of → </div> <div style="text-align: left; margin-left: 10px;"> ← By To obtain </div> </div> | Btu/(hr) (ft ²) (F°) | Btu/(hr) (in. ²) (F°) | Watts/(in. ²) (F°) | Watts/(cm ²) (C°) | Kg-calories/(hr) (cm ²) (C°) | Gram-calories/(sec) (cm ²) (C°) |
|---|----------------------------------|-----------------------------------|--------------------------------|-------------------------------|--|---|
| Btu/(hr) (ft ²) (F°) | 1 | 144 | 490 | 1760 | 2050 | 7380 |
| Btu/(hr) (in. ²) (F°) | 6.94 × 10 ⁻³ | 1 | 3.424 | 12.2 | 14.2 | 51.3 |
| Watts/(in. ²) (F°) | 2.04 × 10 ⁻³ | 0.292 | 1 | 3.58 | 4.16 | 15.0 |
| Watts/(cm ²) (C°) | 5.68 × 10 ⁻⁴ | 8.19 × 10 ⁻² | 0.279 | 1 | 1.16 | 4.23 |
| Kg-calories/(hr) (cm ²) (C°) | 4.88 × 10 ⁻⁴ | 7.04 × 10 ⁻² | 0.240 | 0.859 | 1 | 3.60 |
| Gram-calories/(sec) (cm ²) (C°) | 1.355 × 10 ⁻⁴ | 1.95 × 10 ⁻² | 6.67 × 10 ⁻² | 0.236 | 0.277 | 1 |

19-3 VIBRATION ISOLATION^(6,8,9,10,11,12)

The necessity of vibration isolation for the proper functioning of any piece of ordnance equipment is undesirable. A thorough search for an alternate device should be made before using a device that requires vibration isolation. Some devices are sufficiently delicate to require isolation from the vibration of the base upon which they will be mounted; other devices produce vibration which must not be transmitted to the base. For such devices, vibration isolation is necessary.

Figure 19-1A shows a typical vibration isolation mounting. The vibration mounts

consist of a piece of rubber bonded to two pieces of metal in such a way that any motion transmitted from one piece of metal to the other must pass through the rubber. One piece of metal is attached to the base, and the other is attached to the device that is to be isolated. The elastic and internal damping properties of rubber make it well-suited for use in vibration isolation devices, although metallic springs, cork, and other materials are also used. Vibration isolation is obtained by making the resonant frequency of the device and its vibration mounts appreciably lower than the forcing frequency which is to be attenuated.

By analysis of the equivalent circuit (Fig. 19-1B) of the vibration isolation mounting (Fig. 19-1A), an expression for the amplitude of vibration of the isolated load is given as

$$y = x \frac{2\zeta \frac{s}{\omega_n} + 1}{\frac{s^2}{\omega_n^2} + 2\zeta \frac{s}{\omega_n} + 1} \quad (19-6)$$

where

y = amplitude of vibration of isolated load in ft

x = amplitude of vibration of base, in ft

$\zeta = \frac{f}{2} \sqrt{\frac{g}{mk}} = \text{damping ratio}$

f = damping of vibration mounts, in lb-sec/ft

$g = 32.2 \text{ ft/sec}^2 = \text{acceleration of gravity}$

m = mass of isolated load, in lb

k = combined spring constant of vibration mounts, in lb/ft

s = Laplace operator

$\omega_n = \sqrt{\frac{gk}{m}} = \text{undamped resonant frequency in radians/sec}$

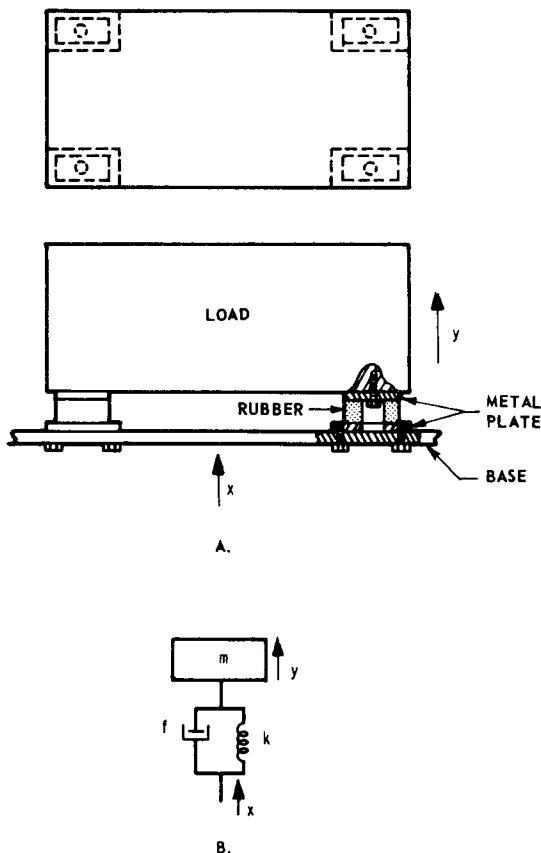


Fig. 19-1 Vibration isolation mounting.

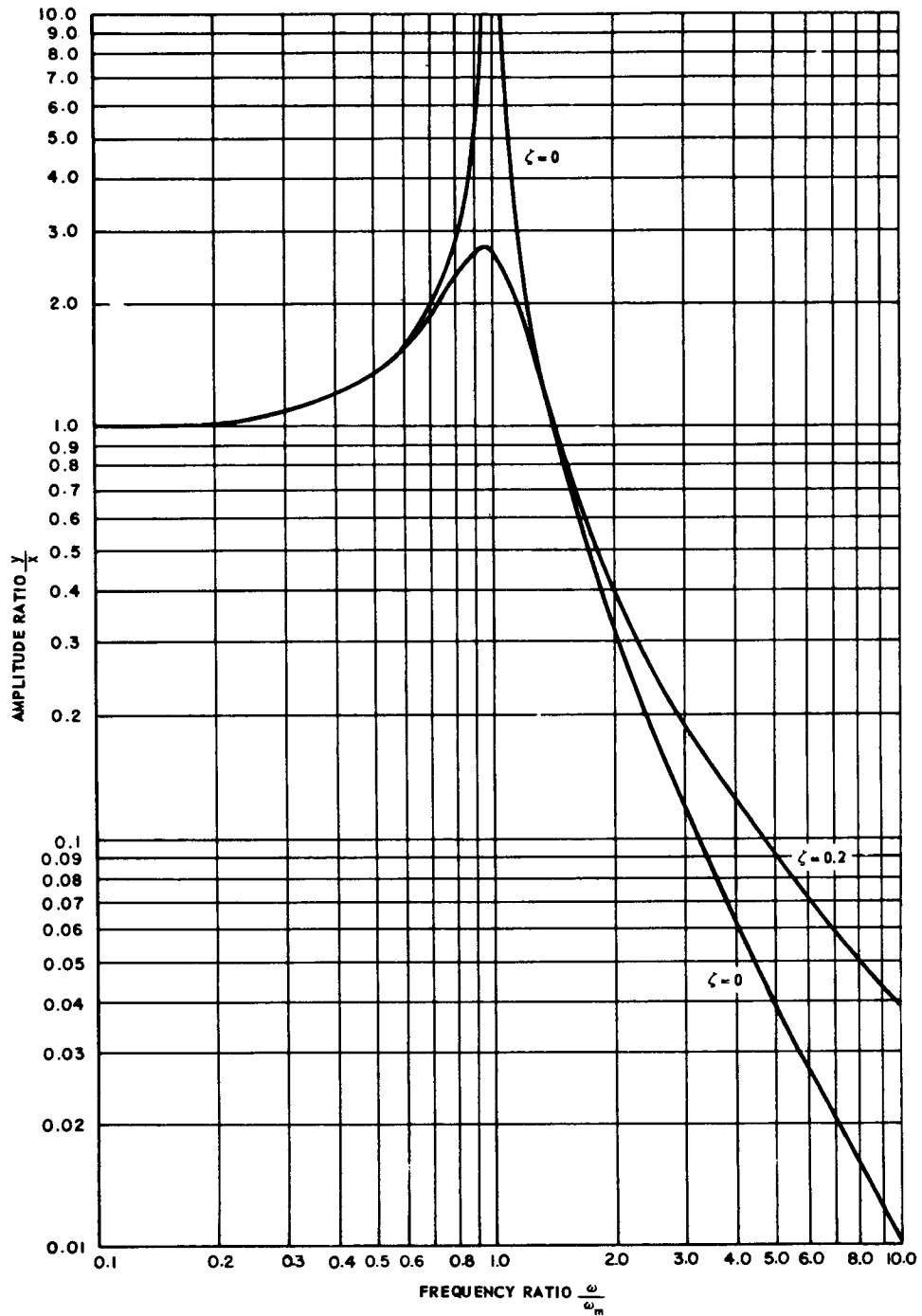


Fig. 19-2 Frequency response of vibration isolation mounting.

The amplitude of the ratio y/x as a function of the ratio of the forcing frequency (ω) to the resonant frequency (ω_n) is plotted in Fig. 19-2 for values of ζ equal to zero and 0.2. It can be seen that, for the damped or undamped condition, extremely large amplitudes occur when the forcing frequency equals the resonant frequency. Most devices equipped with vibration mounts either pass through the mount resonant frequency quickly or are only occasionally excited by a forcing frequency equal to the resonant frequency. Most vibration isolation mountings have a resonant frequency that is less than half the forcing frequency. The resonant frequency must be in the order of one-tenth of the forcing frequency if less than 5 percent of the amplitude of the forcing frequency is to be transmitted through the vibration isolation mounts.

The natural frequency of a mass that is mounted for vibration isolation in the vertical direction is determined by the weight of the load and the static deflection of the mounts when supporting the weight of the load. As the deflection y (ft) produced by a weight w (lb) acting on a spring with a constant k (lb/ft) is

$$y = w/k \quad (19-7)$$

the natural frequency of the mass will be

$$\omega_n = \sqrt{\frac{g}{y}} = \frac{5.67}{\sqrt{y}} \text{ radians/sec} = \frac{0.902}{\sqrt{y}} \text{ cps} \quad (19-8)$$

where y is in feet, or

$$\omega_n = \frac{3.13}{\sqrt{y}} \text{ cps}$$

where y is in inches.

The analysis given here is simplified to include only motion along a line perpendicular to the mounting surface, assuming loads for which the center of mass is equidistant from the isolation supports. In practice, it is necessary to consider rotary vibration of the load, particularly if vibrations can occur parallel to the mounting surface or if rotary forcing vibrations can occur about any axis.

Vibration isolation can be made unnecessary in some cases by eliminating the vibration at its source. Various friction, viscous, and tuned damping devices have been used for this purpose. Static and dynamic balancing are also used to eliminate vibration caused by rotating devices.

19-4 SHOCK ISOLATION^(6,7,10,11,12)

The indiscriminate use of shock isolation devices should be avoided because their use often makes conditions worse instead of better. Therefore, servomechanisms should be equipped with shock mounts only when absolutely necessary. A large variety of electrical and mechanical devices are available which, without shock protection, will withstand shocks higher than those normally encountered under service conditions. Even electron tubes, delicate as they may seem, can often operate satisfactorily without shock protection. If precautions are taken in the design of the control system components, there should be no need for devices to reduce shock

intensity in the majority of ordnance servomechanisms. Table 19-6 lists typical vibration and shock values encountered in ordnance equipment.

The physical construction of mounts for shock isolation is identical with that of mounts for vibration isolation. In general, the underlying basis for both types of mounts is the same in that they serve to lower the acceleration and hence the force applied to the device to be protected. The resonant frequency of loads supported by resilient shock mounts is often set at approximately 20 cps because a resonant frequency of 20 cps provides adequate protection against most

CONSTRUCTIONAL TECHNIQUES

TABLE 19-6 TYPICAL VIBRATION AND SHOCK VALUES ENCOUNTERED IN ORDNANCE EQUIPMENT

| Equipment | Source | Frequency (cps) | Maximum Acceleration in g's† |
|-----------------------------|---|--------------------------------|---|
| Trucks and on-carriage guns | springs | 2-5 | 5-7 |
| | tires | 8-15 | |
| | structures | 60-200 | |
| Tank | ———— | 3 times speed of tank (in mph) | 5 (on black top road) 25 (cross country) 2 (on level gravel road) |
| ———— | muzzle blast 6 ft from equipment | impulse | 1000-1500 (exterior) 300-1250 (interior) |
| ———— | moderate non-penetrating ballistic impact | impulse | 400 |
| ———— | heavy non-penetrating ballistic impact | impulse | vehicle { 1800 (fore & aft) 900 (lateral) 275 (vertical) |
| | | | equipment { 260 (fore & aft) 375 (lateral) |

†Maximum damage due to ballistic impact appears to be caused by accelerations of approximately 3000 g.

Data from letter to Jackson & Moreland, Inc., Boston, Mass., Record No. 5817, from Ordnance Corps, Frankford Arsenal, 7 August 1956.

shocks. Furthermore, most vibrations that might excite the system and cause the spring-mass system to resonate are below 20 cps and hence are not amplified by the shock mounts through sympathetic vibration.

The analytical model of a shock-mounted device is shown in Fig. 19-3. In the figure, k_m is the spring constant and f_m is the damping of the shock mount. The mass of the frame to which the shock mounts are attached to support the load is m_1 . The mass

of the protected device is m_2 , k_d is the spring constant, and f_d is the damping of the support. An example of a shock-mounted device is an instrument mounted in a heavy cabinet which, in turn, is mounted on resilient shock mounts. The amount of protection provided by the shock mounts depends as much upon the dynamic characteristic of the system involving k_d , f_d , and m_2 as it does upon the dynamic characteristics of the shock mounts and the mass m_1 . If it is assumed that m_1

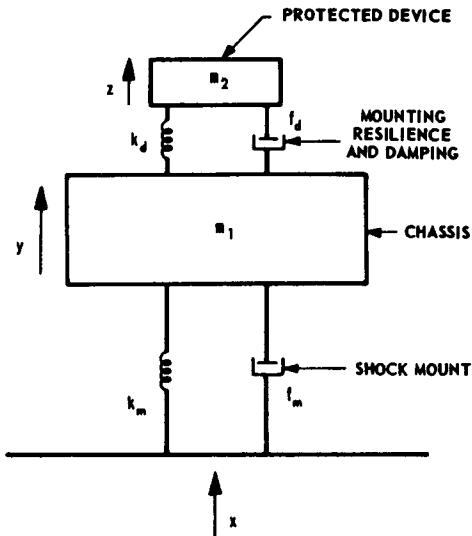


Fig. 19-3 Analytical model of shock-mounted device.

is so large that its motion is unaffected by motion of m_2 , the following general characteristics are true of the system:

(a) To be of any significant value, the

shock mounts must be such that the resonant frequency of m_1-k_m system is less than half the resonant frequency of the m_2-k_d system. This means that the mounting of the protected device on the chassis or frame must be relatively stiff. If the two frequencies coincide, the acceleration of m_2 becomes extremely large for any reasonable value of damping. If the resonant frequency of the m_1-k_m system is much above the resonant frequency of the m_2-k_d system, the resulting acceleration of m_2 is essentially the same whether or not shock mounts are used.

b) The acceleration of m_2 , as well as the acceleration of m_1 , rises when the shock-mount damping f_m is increased. However, some damping is necessary to prevent vibration for excessive lengths of time after a shock occurs.

(c) For a shock consisting of an impulse (which might be represented by a sine wave with a time duration of a half-cycle) at position x in Fig. 19-3, the acceleration of m_1 is greatest when the shock duration is half the resonant period of the m_1-k_m system, provided there is no damping (f_m). The peak acceleration occurs after the shock impulse has again returned to zero.

BIBLIOGRAPHY

- 1 W. W. Woodson, *Human Engineering Guide for Equipment Designers*, University of California Press, Berkeley, Calif., 1954.
- 2 Tufts College for Applied Experimental Psychology, *Handbook of Human Engineering*, Special Devices Center (U. S. Navy) Human Engineering Project 20-G-1, Technical Report SDC 199-1-1.
- 3 W. H. McAdams, *Heat Transmission*, McGraw-Hill Book Company, Inc., New York, N. Y., 1933.
- 4 E. Griffiths and A. H. Davis, "The Transmission of Heat by Radiation and Convection", *Department of Scientific and Industrial Research, Food Investigation Special Report No. 9*, His Majesty's Stationery Office, London, England, 1931.
- 5 American Society of Heating and Ventilating Engineers, *Heating, Ventilating, and Air Conditioning Guide*, Vol. 34, p. #284, 1956.

CONSTRUCTIONAL TECHNIQUES

- 6 C. E. Crede, *Vibration and Shock Isolation*, John Wiley & Sons, Inc., New York, N. Y., 1951.
- 7 W. W. Soroka, "Shock Isolation", *Product Engineering*, Vol. 26, p. #167, June, 1955.
- 8 H. S. Ryder and E. K. Gatcombe, "Designing Vibration Absorbers", *Machine Design*, Vol. 21, p. #142, November, 1949.
- 9 A. L. Kimball, *Vibration Prevention In Engineering*, John Wiley & Sons, Inc., New York, N. Y., 1932.
- 10 J. Markowitz, "Shock or Vibration Isolators", *Product Engineering*, Vol. 24, p. #212, June, 1953.
- 11 Den Hartog, *Mechanical Vibrations*, McGraw-Hill Book Company, Inc., New York, N. Y., 1947.
- 12 S. Timoshenko, *Vibration Problems in Engineering*, D. Van Nostrand Co., Inc., New York, N. Y., 1937.

CHAPTER 20

SUPPLEMENTARY TABLES, FORMULAS, AND CHARTS*

20-1 MASS MOMENT OF INERTIA

20-1.1 DEFINITION

The mass moment of inertia of a body with respect to any axis is defined as

$$J = \int r^2 dm \quad (20-1)$$

where J is the moment of inertia of the body, dm is an element of mass of the body, and r is the perpendicular distance of dm from the axis about which the inertia is to be found. The integral must be evaluated over the entire volume of the body. The dimensions of mass moment of inertia are ML^2 . The relations between the commonly used units of mass moment of inertia are listed in Table 20-1.

20-1.2 DATA FOR EQUATIONS

Table 20-2 is included to facilitate writing equations involving torque, moment of inertia, and acceleration. When writing equations using torque in units of the first column of the table and moment of inertia in units of the second column, divide the moment of inertia by the number in the third column of the table to obtain acceleration in terms of radians per second per second.

20-1.3 PARALLEL-AXIS THEOREM

If the moment of inertia of a body about an axis through its center of mass is known, and if it is desired to obtain the moment of inertia about an axis parallel to that particular centroidal axis, then

$$I = I_o + Mr^2 \quad (20-2)$$

where

I = moment of inertia about the new axis

I_o = moment of inertia about the centroidal axis

M = mass of the body

r = distance between centroidal axis and the new axis

Equation (20-2) is called the *parallel-axis theorem* and is proven in most kinematics tests.

20-1.4 PRINCIPAL AXES OF INERTIA

For any particular point within a body, there is one axis about which the moment of inertia is a maximum and another axis about which the moment of inertia is a minimum. These two axes are perpendicular to each other. These axes, together with the axis through their point of intersection and perpendicular to the plane formed by them, are called the *principal axes of inertia* of the body.

20-1.5 PRODUCT OF INERTIA

The product of inertia of a body with respect to two mutually perpendicular planes is defined as

$$G_{ab} = \int ab \, dm \quad (20-3)$$

where a and b are the perpendicular distances from mass element dm to the two planes. A body must have three products of inertia

$$G_{xy} = \int xy \, dm, G_{yz} = \int yz \, dm, G_{zx} = \int xz \, dm \quad (20-4)$$

*By J. O. Silvey

POWER ELEMENTS AND SYSTEM DESIGN

TABLE 20-1 RELATIONS BETWEEN UNITS OF MASS MOMENT OF INERTIA

| Multiply number of —→ To obtain ↓ By ↘ | | gram-cm ² | oz-in. ² | lb-in. ² | lb-ft ² | slug-ft ² |
|---|--|------------------------|------------------------|------------------------|------------------------|----------------------|
| gram-cm ² | | 1 | 182.9 | 2926 | 4.214×10^{-5} | 1.356×10^7 |
| oz-in. ² | | 5.467×10^{-3} | 1 | 16 | 2304 | 7.412×10^4 |
| lb-in. ² | | 3.418×10^{-4} | 0.0625 | 1 | 144 | 4632 |
| lb-ft ² | | 2.373×10^{-6} | 4.340×10^{-4} | 6.944×10^{-3} | 1 | 32.17 |
| slug-ft ² | | 7.375×10^{-8} | 1.349×10^{-5} | 2.159×10^{-4} | 0.03108 | 1 |

**TABLE 20-2 RELATIONS BETWEEN TORQUE, MOMENT OF INERTIA,
AND ACCELERATION**

| A torque of | Will accelerate a body having a moment of inertia of | To |
|-------------|---|-------|
| 1 dyne-cm | 1 gram-cm ² | 1 |
| 1 gram-cm | 1 gram-cm ² | 980.7 |
| 1 oz-in. | 1 oz-in. ² | 386 |
| 1 lb-in. | 1 lb-in. ² | 386 |
| 1 lb-ft | 1 lb-ft ² | 32.17 |
| 1 lb-ft | 1 slug-ft ² | 1 |

SUPPLEMENTARY TABLES, FORMULAS, AND CHARTS

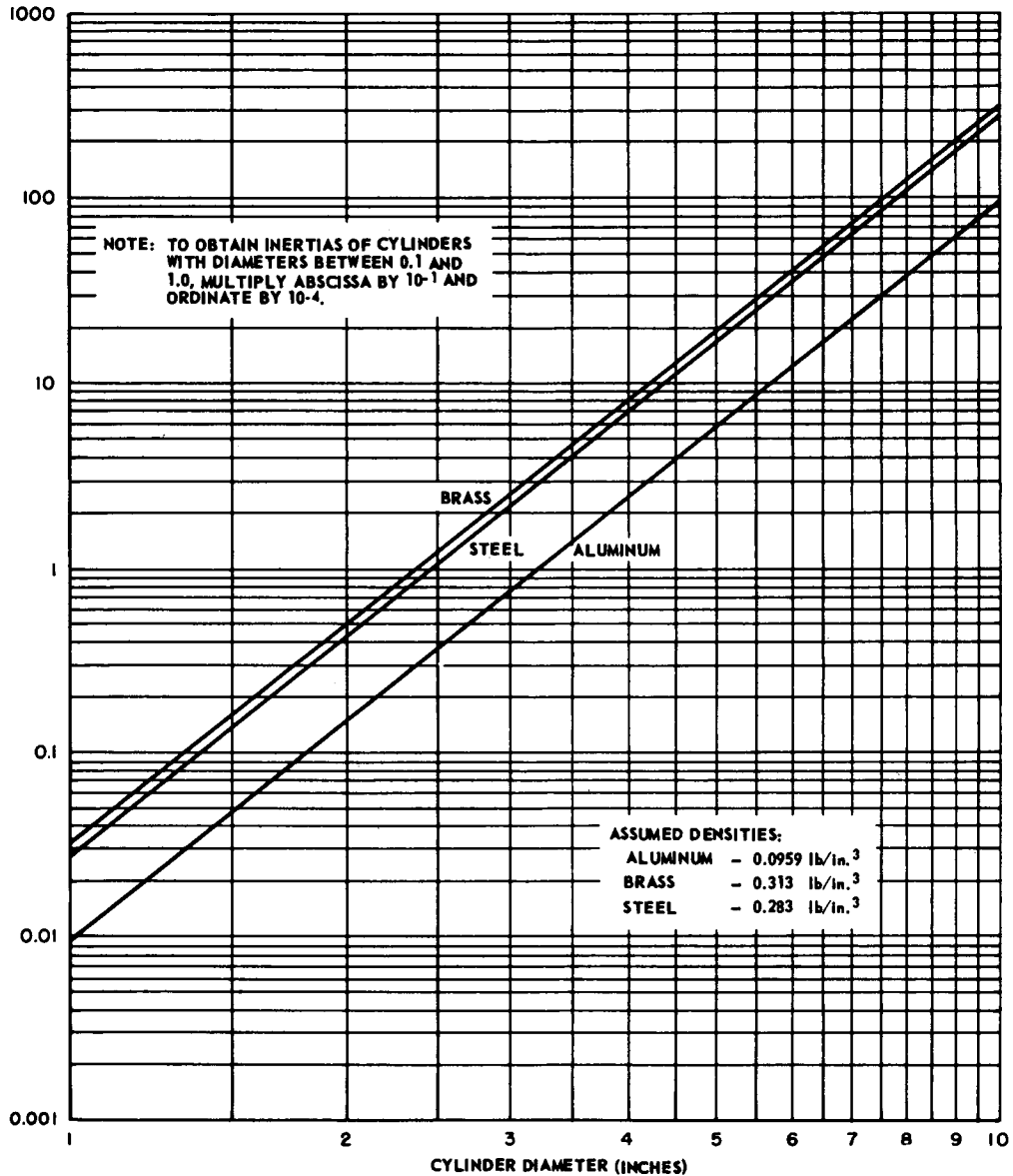


Fig. 20-1 Moments of inertia about the principal axes of cylinders one inch long.

The product of inertia of a body with respect to any two of the planes formed by the principal axes of the body is zero.

20-1.6 INERTIA WITH RESPECT TO A LINE THROUGH THE ORIGIN

The moment of inertia of a body with respect to any line passing through the coordinate origin is

$$\begin{aligned} J &= J_x \cos^2 \alpha + J_y \cos^2 \beta + J_z \cos^2 \gamma \\ &\quad - 2G_{xy} \cos \alpha \cos \beta - 2G_{yz} \cos \beta \cos \gamma \\ &\quad - 2G_{zx} \cos \gamma \cos \alpha \end{aligned} \quad (20-5)$$

where

J_x, J_y, J_z = moments of inertia about the x -, y -, and z -axes, respectively

α, β, γ = angles between the x -, y -, and z -axes, respectively

G_{xy}, G_{yz}, G_{zx} = products of inertia with respect to the yz - and xz -planes, the xz - and xy -planes, and the xy - and yz -planes, respectively.

If the coordinate system is so chosen that the x -, y -, and z -axes lie along the principal axes of the body, the products of inertia are zero and Eq. (20-5) reduces to

$$J = J_x \cos^2 \alpha + J_y \cos^2 \beta + J_z \cos^2 \gamma \quad (20-6)$$

20-1.7 TABULATED MOMENTS OF INERTIA

Expressions for the moments of inertia of a large number of bodies are tabulated in the Handbook of Engineering Fundamentals⁽¹⁾ and the Handbook of Chemistry and Physics.⁽²⁾ The moments of inertia of cylinders frequently used in control work are plotted in Fig. 20-1.

20-1.8 COMPLICATED SHAPES

The moment of inertia of a body with a complicated shape can be approximated by dividing the body into a number of parts having simple shapes, computing the inertia of each part, and then adding the inertias of the individual parts. If the shapes of the parts can be approximated by simple geometric shapes, the tabulated expressions for moments of inertia and the parallel-axis theorem can be used to compute the total moment of inertia of the body. If the body is extremely irregular, it sometimes is more convenient to plot its cross section on graph paper, using either Cartesian or polar coordinates (depending upon the problem), and to compute the moment of inertia of the rectangles or arcs covered by the figure. For bodies such as tank turrets, the moment of inertia must be computed at several elevations, and the inertias of the various cross sections must be added to obtain the total inertia.

20-2 DAMPING AND FRICTION

20-2.1 VISCOSITY

20-2.2 Definition

Viscosity is that characteristic of a fluid which produces a resistive force when there is relative motion between a solid body and a liquid with which it is in contact. This resistive force, proportional to relative velocity between the solid and the fluid, is known as *viscous friction* and produces damping. The viscous properties of fluids produce damping whenever relative motion between a fluid and

a solid body with which the fluid is in contact occurs. Dampers are made that consist of two parallel discs or concentric cylinders which are separated by a small gap filled with viscous fluid and which are capable of moving relative to one another. Viscous damping also occurs when a viscous fluid flows in pipes or tubing.

20-2.3 Absolute (Dynamic) Viscosity

If the space between two large parallel plates of area a , separated by a distance l , is

SUPPLEMENTARY TABLES, FORMULAS, AND CHARTS

filled with a fluid of viscosity μ , and if one plate is moved relative to the other with velocity v , the separation being kept constant, then the force f required to sustain the motion of the moving plate is

$$f = \frac{\mu av}{l} \quad (20-7)$$

From Eq. (20-7), the viscosity is $\mu = fl/av$ and has the dimensions $ML^{-1}T^{-1}$ or $FL^{-2}T$. Viscosity defined in this way is called *absolute*, or *dynamic, viscosity*. The only unit of absolute viscosity that has a name is the poise, which is one dyne-second per square centimeter, or one gram mass per centimeter-second. Table 20-3 lists the relations between various units of absolute viscosity.

20-2.4 Kinematic Viscosity

Kinematic viscosity is defined as the absolute viscosity of a fluid divided by its density. When a dimension system of force, distance, and time is used, the dimensions of kinematic viscosity are L^2T^{-1} . The centimeter-gram-second (cgs) unit is the stoke, which has the dimensions centimeter squared per second. The English unit, which would be used with the pound force-foot-second system commonly employed in engineering work, is the foot squared per second. The relations between the three commonly used units of kinematic viscosity are:

- (a) stokes $\times 100$ = centistokes
- (b) $(ft^2/sec) \times 929$ = stokes

TABLE 20-3 RELATIONS BETWEEN UNITS OF ABSOLUTE VISCOSITY

| <div style="text-align: center;"> </div> | poise* (dyne-seconds per square centimeter) | gram force-seconds per square centimeter | ounce force-seconds per square inch | pound force-seconds per square inch | pound force-seconds per square foot |
|---|---|---|--|--|--|
| poise* (dyne-seconds per square centimeter) | 1 | 980.7 | 4309 | 6.89×10^4 | 479 |
| gram force-seconds per square centimeter | 1.020×10^{-3} | 1 | 28.35 | 454 | 3.15 |
| ounce force-seconds per square inch | 2.33×10^{-4} | 0.03527 | 1 | 16 | 0.111 |
| pound force-seconds per square inch | 1.451×10^{-5} | 2.202×10^{-3} | 0.0625 | 1 | 6.94×10^{-3} |
| pound force-seconds per square foot | 2.09×10^{-3} | 0.317 | 9 | 144 | 1 |

*To obtain centipoises, multiply number of poises by 100.

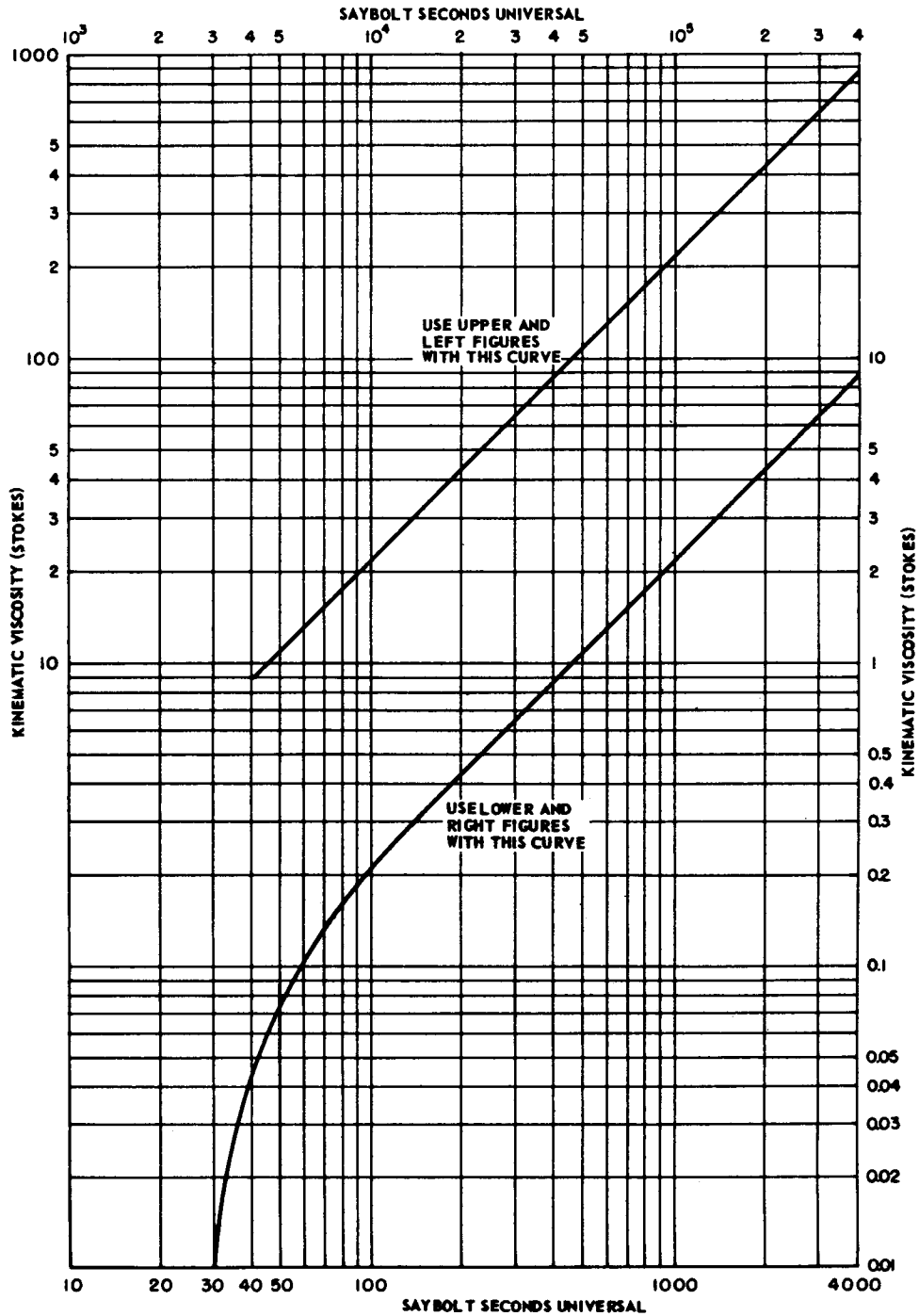


Fig. 20-2 Relation between Saybolt seconds universal and stokes.

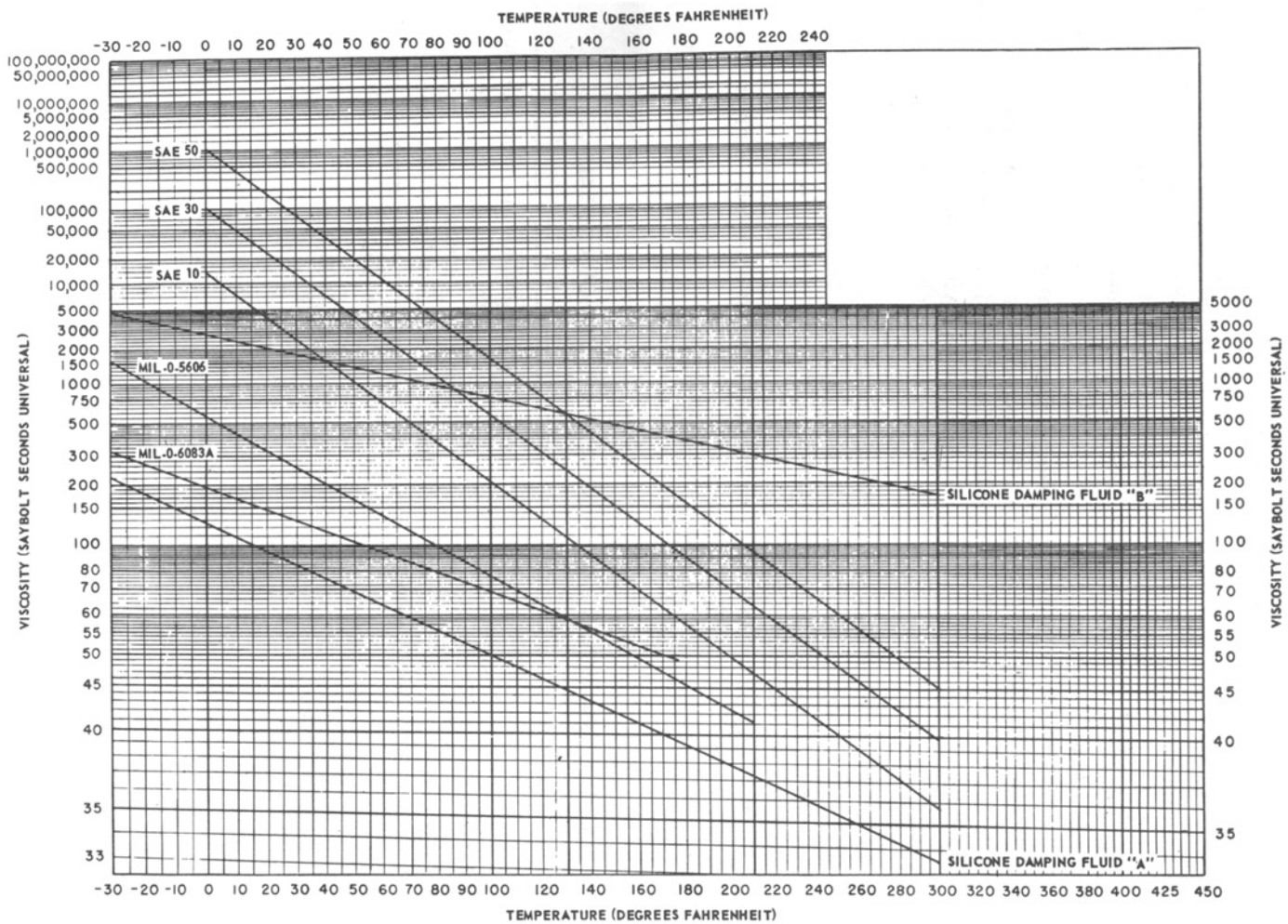


Fig. 20-3 Viscosity versus temperature for some oils and fluids.

The time required for a fluid to drain from a viscometer by the action of gravity is a function of the kinematic viscosity of the fluid. In the United States, the kinematic viscosity of lubricating oils and hydraulic fluids is usually expressed in Saybolt seconds universal (ssu). This is the time, in seconds, for a particular quantity of fluid to escape from a specified vessel through a specified capillary tube. The conversion between the time of discharge and the kinematic viscosity of the fluid, in units suitable for calculations, has been determined empirically and is shown in Fig. 20-2. The conversion for various viscometers used in Europe is given in reference 1.

20-2.5 Effect of Temperature

The viscosity of all lubricating oils, hydraulic fluids, and damping fluids increases when their temperature is lowered. However, most hydraulic fluids contain additives that reduce the change in viscosity below that ex-

hibited by pure petroleum-base oils. The relation between temperature and viscosity for some oils and fluids is shown in Fig. 20-3.

20-2.6 FRICTION

Friction that is proportional to velocity is called *viscous friction* and is usually associated with fluids. Friction that is constant regardless of velocity, but which always opposes motion, is called *coulomb friction* and is associated with poorly lubricated sliding surfaces.

20-2.7 Coefficient of Friction

The coefficient of friction between two sliding surfaces is defined as the tangential friction force divided by the normal force. When no motion occurs, the quotient is called the *coefficient of static friction*. The value of the coefficient of static friction is determined by the maximum friction force that will not produce slipping. When one surface slides relative to the other, the quotient is called the *coefficient of kinetic friction*. Both coefficients

TABLE 20-4 COEFFICIENTS OF FRICTION FOR VARIOUS METALS

| Material | Condition | Coefficient of Kinetic Friction | Coefficient of Static Friction |
|------------------------------|------------------|---------------------------------|--------------------------------|
| Cast iron on cast iron | dry | 0.152 | 0.162 |
| Cast iron on cast iron | greased | 0.08-0.10 | 0.16 |
| Cast iron on bronze | dry | 0.213 | |
| Cast iron on bronze | greased | 0.132 | |
| Wrought iron on wrought iron | dry | 0.44 | |
| Wrought iron on wrought iron | greased | 0.08-0.10 | 0.11 |
| Wrought iron on bronze | dry | 0.18 | 0.19 |
| Wrought iron on bronze | greased | | 0.07-0.08 |
| Steel on steel | dry (10 ft/sec) | 0.09 | 0.15 |
| Steel on steel | dry (100 ft/sec) | 0.03 | |
| Steel on bronze | dry | 0.152 | |

of friction are independent of pressure on the surfaces, and hence independent of the surface area, when the pressure is between 3/4 psi and 100 psi. For pressures below 3/4 psi, the coefficient of friction increases somewhat. It also increases for extremely high pressures.

20-2.8 Characteristics of Coefficient of Friction

The coefficient of static friction is always higher than the coefficient of kinetic friction. The coefficient of kinetic friction is independent of velocity for average relative velocities.

However, the coefficient of kinetic friction decreases at extremely high velocities and increases at extremely low velocities. Since the coefficient of friction, either static or kinetic, is dependent upon the condition of the surfaces and the presence or absence of small amounts of lubricants, any quoted value should be considered as an approximation. The coefficients of friction for various metals are listed in Table 20-4.

20-3 SPRINGS

20-3.1 HELICAL SPRINGS

20-3.2 Stress

Neglecting end effects and curvature correction and assuming round wires, the stress in loose-wound tension helical springs or compression helical springs is

$$S = \frac{8WD}{\pi d^3} \quad (20-8)$$

where

S = maximum fibre stress, in psi

W = load, in lb

D = mean coil diameter, in in.

d = wire diameter in in.

Equation (20-8) results in an error of more than ten percent if the spring index (D/d) is less than 13. Neglecting end effects, assuming round wire, and taking curvature correction into account

$$S = \frac{8WD}{\pi d^3} K \quad (20-9)$$

where S , W , D , and d are defined in Eq. (20-8) and

$$K = \frac{4(D/d) - 1}{4(D/d) - 4} + \frac{0.615}{D/d}$$

K is known as the Wahl correction factor.⁽⁹⁾

20-3.3 Deflection

The deflection of a helical spring, neglecting end effects, is

$$Y = \frac{8WD^3n}{E_s d^4} \quad (20-10)$$

where W , d , and D are defined in Eq. (20-8) and

n = number of active coils

E_s = torsional modulus of rigidity, in psi

Y = spring deflection, in in.

20-3.4 Torsional Elasticity

The torsional elastic properties of some spring materials are listed in Table 20-5. The values of working stress in this table are quite conservative and, in most cases, can be exceeded without over-stressing the spring.

20-3.5 Design Table

Table 20-6 can be used to choose a helical spring size quickly if the desired spring constant is known. The spring constant in pounds per inch of deflection is the spring scale (lower entry) of the table divided by the number of turns in the spring to be used. The upper entry of the table is the spring load that will produce a stress of 100,000 psi, and is independent of the number of turns in the spring. Loads corresponding to higher or lower stresses are

equal to the tabulated loads multiplied by the desired stress in psi and divided by 100,000. Linear interpolation to obtain the spring scale and load for any spring coil diameter or any wire size will result in an error of less than 5 percent; obviously, double interpolation is required if the desired coil diameter and wire diameter both fall between values entered in the table. It should be emphasized that this table is valid only for round wire with a torsional modulus of 11×10^6 psi. An example illustrating the use of this table follows: Find a steel spring with 0.5-inch outside diameter which will deflect 0.5 inch under a load of 7 lb, and which can be deflected 1.0 inch without exceeding 100,000 psi stress. The maximum load will be $7 \left(\frac{1.0}{0.5} \right) = 14$ lb. From the table,

a wire diameter of 0.059 will produce a spring of 0.5-inch outside diameter which will carry 15.3 lb., which is adequate for this application. From the table, the spring scale for a single coil of this spring will be 204 lb/inch. The desired spring constant is $\frac{7}{0.5} = 14$ lb/inch. The

required number of coils is therefore $\frac{204}{14} = 14.6$ coils. Minimum length of a tension spring having the desired characteristics, neglecting length of spring ends, is $(14.6)(.059) = 0.8614$ inch. A compression spring having the desired characteristics would be one inch longer to accommodate the required deflection of one inch before it is compressed solid. Table 20-6 is a condensed version of the table given by Ross.⁽⁵⁾ Other spring design tables appear in the literature,^(6,7,8) as well as in standard mechanical engineers' handbooks.

TABLE 20-5 TORSIONAL ELASTIC PROPERTIES OF SOME SPRING MATERIALS

| Material | Elastic Limit* (psi) | Recommended Working Stress‡ (psi) | Modulus Rigidity (psi) |
|--|----------------------|-----------------------------------|------------------------|
| Music wire | 130,000 | 70,000 | 11×10^6 |
| Hot rolled steel (SAE 1095) hardened to Rockwell C43 | 92,000 | 60,000 | 11×10^6 |
| Stainless steel (Type 302) | 95,000 | 35,000 | 11×10^6 |
| Phosphor or silicon bronze | 67,000 | 30,000 | 6.2×10^6 |
| Beryllium copper | 80,000 | 50,000 | 6.5×10^6 |

*An average value of elastic limit is given here. Actual values vary widely.

‡The recommended working stress given here is approximately 80% of the minimum elastic limit of the material.

20-3.6 CANTILEVER SPRINGS

20-3.7 Definition

A cantilever spring (Fig. 20-4) is made by clamping one end of a bar and applying a load at the other end. In effect, the spring is essentially a cantilever beam and ordinary beam theory can be used to obtain stress and deflection.

20-3.8 Stress and Deflection

The stress of a cantilever beam is

$$S = \frac{MC}{I} \quad (20-11)$$

where

S = maximum tensile fibre stress, in psi

M = moment on beam, in in.-lb

C = distance from neutral axis to outer fibre, in in.

I = moment of inertia of cross section of the beam about the neutral axis, in in.⁴

For a cantilever beam with a rectangular cross section along its entire length (Fig. 20-4), the stress is

$$S = \frac{6Wl}{bh^2} \quad (20-12)$$

where

S = maximum stress, in psi

W = load, in lb

l = distance between clamp and point of application of load, in in.

b and h = cross-sectional dimensions, in in.

The deflection of a cantilever beam under a load W is

$$Y = \frac{Wl^3}{3EI} \quad (20-13)$$

where I , W , and l are defined in Eqs. (20-11) and (20-12), and

Y = beam deflection, in in.

E = Young's modulus for the beam material, in psi

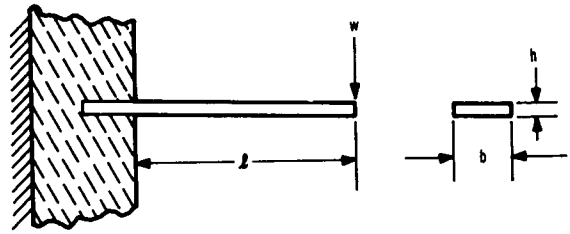


Fig. 20-4 Cantilever spring.

For a rectangular beam (Fig. 20-4), the deflection becomes

$$Y = \frac{4Wl^3}{Ebh^3} = \frac{2Sl^2}{3Eh} \quad (20-14)$$

20-3.9 Tensile Elastic Properties

A cantilever beam bends instead of twisting. Therefore, the modulus of elasticity and the allowable stresses are not the same as for a helical spring. Table 20-7 lists the tensile elastic properties of some spring materials.

20-3.10 TORSION BAR SPRINGS

20-3.11 Definition

Torsion bar springs have one end clamped while a torque is applied to the other end.

20-3.12 Stress and Torque

The equation for stress in a torsion bar spring, assuming a round bar of uniform diameter, is

$$S = \frac{E_s r \theta}{l} \quad (20-15)$$

where

S = shearing stress, in psi

E_s = torsional modulus of rigidity, in psi

r = bar radius, in in.

θ = angle of twist, in radians

l = length of spring, in in.

TABLE 20-6 PERMISSIBLE LOAD AND SPRING SCALE OF HELICAL COMPRESSION SPRINGS

| Wire Diameter (in.) | Outside Diameter of Coil (in.) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------------------------|-----------------------------------|------|------|-----|--|-----|------|-----|---------------|-----|-----|---|---------------|-------|-------|---|---------------|---|-------|---|--------------|--|--|--|--------------|--|--|--|--------------|--|--|--|--------------|--|--|--|--------------|--|--|--|--------------|--|--|--|--------------|--|--|--|--------------|--|--|--|
| | 1/8 | 5/32 | 3/16 | 1/4 | 5/16 | 3/8 | 7/16 | 1/2 | 5/8 | 3/4 | 7/8 | 1 | 1-1/4 | 1-1/2 | 1-3/4 | 2 | 2-1/2 | 3 | 3-1/2 | 4 | 4-1/2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.010 | 0.305 9.47 | | | | Load in lb. at 100,000 psi — Wahl's Factor (K) included Spring scale per single turn, in lb/in. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.012 | 0.522 20.7 | | | | 0.421 10.0 | | | | 0.350 5.43 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.014 | 0.823 40.3 | | | | 0.669 19.4 | | | | 0.560 10.5 | | | | 0.422 4.18 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.016 | 1.21 72.9 | | | | 0.983 34.4 | | | | 0.824 18.4 | | | | 0.626 7.37 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.018 | 1.71 124 | | | | 1.040 57.4 | | | | 1.17 30.7 | | | | 0.889 12.1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.020 | 2.32 200 | | | | 1.89 91.3 | | | | 1.60 48.6 | | | | 1.22 18.9 | | | | 0.972 9.06 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.024 | 3.90 464 | | | | 3.22 208 | | | | 2.72 108 | | | | 2.09 41.4 | | | | 1.68 19.9 | | | | 1.41 11.1 | | | | 1.21 6.67 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.029 | 6.61 1160 | | | | 5.55 496 | | | | 4.71 253 | | | | 3.64 94.1 | | | | 2.94 44.6 | | | | 2.50 24.6 | | | | 2.13 14.9 | | | | 1.87 9.75 | | | | | | | | | | | | | | | | | | | | | | | |
| 0.033 | | | | | 7.97 916 | | | | 6.85 460 | | | | 5.29 166 | | | | 4.31 77.5 | | | | 3.61 42.4 | | | | 3.11 25.6 | | | | 2.74 16.7 | | | | 2.21 8.26 | | | | | | | | | | | | | | | | | | | |
| 0.037 | | | | | 10.9 1580 | | | | 9.42 785 | | | | 7.41 280 | | | | 6.01 128 | | | | 5.07 69.6 | | | | 4.41 42.2 | | | | 3.88 27.3 | | | | 3.10 13.3 | | | | 2.61 7.43 | | | | | | | | | | | | | | | |
| 0.043 | | | | | | | | | 14.4 1618 | | | | 11.4 559 | | | | 9.33 249 | | | | 7.91 134 | | | | 6.82 80.0 | | | | 6.07 51.6 | | | | 4.89 25.0 | | | | 4.10 14.0 | | | | 3.52 8.60 | | | | | | | | | | | |
| 0.051 | | | | | | | | | 18.5 1233 | | | | | | | | 15.3 543 | | | | 13.0 288 | | | | 11.3 168 | | | | 10.0 107 | | | | 8.01 51.1 | | | | 6.78 28.6 | | | | 5.79 17.3 | | | | 5.14 11.4 | | | | | | | |
| 0.059 | | | | | | | | | 27.6 2509 | | | | | | | | 23.2 1059 | | | | 19.9 554 | | | | 17.3 322 | | | | 15.3 204 | | | | 12.4 96.3 | | | | 10.4 52.7 | | | | 9.01 32.2 | | | | 7.89 21.1 | | | | 6.33 10.3 | | | |
| 0.071 | | | | | | | | | | | | | 38.7 2563 | | | | 33.4 1295 | | | | 29.4 741 | | | | 26.1 463 | | | | 21.3 215 | | | | 18.0 117 | | | | 15.5 70.6 | | | | 13.6 45.3 | | | | 11.1 22.4 | | | | 9.21 12.6 | | | |
| 0.085 | | | | | | | | | | | | | | | | | 55.5 3083 | | | | 49.2 1708 | | | | 44.1 1053 | | | | 36.1 478 | | | | 30.5 255 | | | | 26.6 153 | | | | 23.3 97.5 | | | | 18.8 47.4 | | | | 15.8 26.6 | | | |

| Wire Diameter (in.) | Outside Diameter of Coil (in.) | | | | | | | | | | | | | | | | | | | | | |
|------------------------|-----------------------------------|------|------|-----|--------------|--------------|--------------|-----|--------------|--------------|-------------|--------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|--------------|--------------|
| | 1/8 | 5/32 | 3/16 | 1/4 | 5/16 | 3/8 | 7/16 | 1/2 | 5/8 | 3/4 | 7/8 | 1 | 1-1/4 | 1-1/2 | 1-3/4 | 2 | 2-1/2 | 3 | 3-1/2 | 4 | 4-1/2 | |
| 0.092 | | | | | 69.3 4530 | 61.0 2490 | 55.1 1515 | | 45.6 683 | 38.4 363 | 33.4 215 | 29.6 136 | 23.8 66.4 | 19.9 36.8 | 17.1 22.6 | | | | | | | |
| 0.106 | | | | | | 91.0 4973 | 81.9 2967 | | 68.3 1298 | 58.1 679 | 50.8 400 | 44.8 255 | 36.2 121 | 30.1 66.5 | 26.1 40.9 | 22.8 26.6 | | | | | | |
| 0.125 | | | | | | | 130 6710 | | 109 2817 | 93.8 1436 | 81.9 832 | 72.5 525 | 58.8 246 | 49.4 135 | 42.06 82.0 | 37.6 53.3 | 30.2 26.2 | | | | | |
| 0.148 | | | | | | | | | | 151 3166 | 133 1802 | 119 1119 | 96.3 514 | 81.2 279 | 70.5 168 | 62.0 109 | 49.7 53.1 | 41.7 29.7 | | | | |
| 0.177 | | | | | | | | | | 250 7485 | 222 4150 | 198 2526 | 163 1145 | 137 609 | 119 361 | 105 234 | 84.7 113 | 71.0 63.2 | 61.2 38.4 | | | |
| 0.207 | | | | | | | | | | | 341 8857 | 307 5266 | 255 2320 | 217 1218 | 188 716 | 166 457 | 134 218 | 113 121 | 97.2 74.1 | 85.2 48.5 | 76.0 33.5 | |
| 0.244 | | | | | | | | | | | | 484 11805 | 406 4988 | 349 2562 | 306 1488 | 269 941 | 218 443 | 183 243 | 158 148 | 139 95.9 | 125 66.7 | |
| 0.283 | | | | | | | | | | | | | 614 10199 | 534 5154 | 467 2919 | 415 1820 | 338 849 | 285 460 | 247 276 | 216 180 | 194 123 | |
| 0.331 | | | | | | | | | | | | | | 824 10785 | 733 6048 | 652 3713 | 534 1693 | 453 911 | 391 543 | 343 348 | 308 238 | |
| 0.375 | | | | | | | | | | | | | | 1166 19966 | 1038 10903 | 932 6652 | 767 2960 | 652 1575 | 567 933 | 497 597 | 445 406 | |
| 0.4375 | | | | | | | | | | | | | | | 1587 23202 | 1433 13859 | 1201 6005 | 1019 3132 | 888 1835 | 783 1166 | 704 785 | |
| 0.500 | | | | | | | | | | | | | | | | 2073 26645 | 1742 11268 | 1500 5736 | 1310 3330 | 1160 2097 | 1041 1405 | |
| 0.5625 | | | | | | | | | | | | | | | | | | 2426 19853 | 2096 9957 | 1832 5686 | 1628 3546 | 1468 2359 |
| 0.625 | | | | | | | | | | | | | | | | | | 3230 33289 | 2825 16386 | 2491 9240 | 2221 5710 | 1997 3773 |

Adapted by permission from *Machine Design*, Volume 19, No. 3, March, 1947, from article entitled "Helical Spring Design Tables", by H. F. Ross.

The torque is

$$T = \frac{E_s J \theta}{l} = \frac{\pi E_s r^4 \theta}{2l} \quad (20-16)$$

where E_s , θ , and l are defined in Eq. (20-15) and J is the polar moment of inertia of the cross section of the bar about its geometric center, in in.⁴. The torsional elastic properties of some spring materials are listed in Table 20-5. For discussions of Bellville spring washers, helical springs used as torsion springs, and clock or spiral springs, see Wahl.⁽⁴⁾

20-3.13 VIBRATION IN SPRINGS

All springs can be considered as systems of distributed mass, elastance, and damping. In most springs, the damping is extremely small. Therefore, they will resonate at a large number of frequencies with different modes of vibration. Because the damping is low, the magnitude of the spring motion may become extremely high. In most control applications, however, the resonant frequency of springs is above the system forcing frequency, and

resonance in the spring itself is of little interest.

20-3.14 Natural Frequency

The natural frequency of a helical spring with both ends fixed is⁽⁹⁾

$$f = \frac{d}{2\pi r^2 n} \sqrt{\frac{E_s g}{32\rho}} \quad (20-17)$$

where

d = spring wire diameter, in in.

ρ = density of spring material, in lb/in.³

g = acceleration of gravity, in in./sec²

E_s = torsional modulus of rigidity, in psi

r = mean radius of spring coil, in in.

n = number of active turns

f = lowest natural frequency of the spring, in cps

For a steel spring

$E_s = 11.5 \times 10^6$ psi

$\rho = 0.285$ lb/in.³

TABLE 20-7 TENSILE ELASTIC PROPERTIES OF SOME SPRING MATERIALS

| Material | Elastic Limit* (psi) | Recommended Working Stress‡ (psi) | Modulus of Elasticity (psi) |
|---|-------------------------|--------------------------------------|--------------------------------|
| Clock spring steel | 200,000 | 100,000 | 32×10^6 |
| Music wire | 200,000 | 110,000 | 30×10^6 |
| Hot rolled steel (SAE 1095) hardened to Rockwell C43 | 120,000 | 80,000 | 28×10^6 |
| Stainless steel (Type 302) | 150,000 | 40,000 | 28×10^6 |
| Phosphor or silicon bronze | 80,000 | 40,000 | 15×10^6 |
| Beryllium copper | 130,000 | 75,000 | 17×10^6 |

*An average value of elastic limit is given here. Actual values vary widely.

‡The recommended working stress given here is approximately 80% of the minimum elastic limit of the material

Thus, the lowest natural frequency of the steel spring is

$$f = \frac{3510d}{r^2n} \text{ cps} \quad (20-18)$$

There will also be natural frequencies at 2, 3, 4, etc. times the lowest natural frequency. The natural frequency of helical compression springs is also discussed by Maier⁽⁸⁾.

20-4 MISCELLANEOUS CONSTANTS OF FLUIDS AND TUBING USED IN HYDRAULIC SYSTEMS

20-4.1 BULK MODULUS

The bulk modulus of a liquid is a measure of the amount the liquid is compressed when the pressure applied to the liquid is increased. Bulk modulus can be defined as

$$B = \frac{\Delta P}{\left(\frac{\Delta V}{V}\right)} \quad (20-19)$$

where

B = bulk modulus

ΔP = change in pressure

ΔV = change in volume

V = volume of liquid to which pressure is applied

Bulk modulus has the dimensions of FL^{-2} and is usually expressed in lb/in.², although normal atmospheric pressure is sometimes used as the basic unit. Although bulk modulus changes somewhat with pressure, it is essentially constant for petroleum-base hydraulic fluid at pressures between 15 and 3000 psi. A good average value is 2.5×10^5 psi.

20-4.2 COMPRESSIBILITY

Compressibility is the reciprocal of bulk

modulus. For most hydraulic fluids, it is approximately 4×10^{-6} in.²/lb.

20-4.3 SPECIFIC GRAVITY

The specific gravity of most petroleum products is expressed in terms of degrees Baumé or degrees API. These units are the readings of particular hydrometers. Conversion factors between API degrees and specific gravity are given in most handbooks for mechanical engineers. Most hydraulic fluids have a specific gravity of between 0.85 and 0.90 (density between 0.0307 and 0.0325 lb/in.³) at a temperature of 60°F.

20-4.4 ELASTIC PROPERTIES OF TUBING

The elastic properties of the materials most commonly used for tubing in hydraulic systems are listed in Table 20-8. In general, the factor of safety used should depend upon the nature of the driven load. For copper and aluminum, which have no definite yield point, a factor of safety below 2 is never used. It is usually kept above 4. Heavy pulsating loads require an even higher factor of safety. The tensile modulus of elasticity is included in Table 20-8 for use in computing the increased volume of tubing when the tubing is subjected to internal pressure.

TABLE 20-8 ELASTIC PROPERTIES OF TUBING MATERIAL

| Material | Yield Stress (psi) | Ultimate Strength (psi) | Tensile Modulus of Elasticity (psi) |
|---------------------------------|--------------------|-------------------------|-------------------------------------|
| Soft-drawn copper* | 7000 | 33,000 | 15×10^6 |
| Hard-drawn copper* | 49,000 | 55,000 | 15×10^6 |
| Soft aluminum* | 3000 | 9000 | 10×10^6 |
| Alloyed, heat-treated aluminum* | 30,000 | 45,000 | 10×10^6 |
| Steel (SAE 1020) | 40,000 | 65,000 | 29×10^6 |
| Stainless steel (AISI 302) | 40,000 | 90,000 | 29×10^6 |

*These materials have no well-defined yield point. Values are for 0.2% offset.

BIBLIOGRAPHY

- 1 O. W. Eshbach (editor), *Handbook of Engineering Fundamentals* (2nd edition), John Wiley & Sons, Inc., New York, N. Y., 1952.
- 2 C. D. Hodgman (editor), *Handbook of Chemistry and Physics* (38th edition), Chemical Rubber Publishing Company, Cleveland, Ohio, 1956.
- 3 R. T. Kent (editor), *Mechanical Engineer's Handbook, Design and Production Volume* (12th edition), John Wiley & Sons, Inc., New York, N. Y., 1950.
- 4 A. M. Wahl, "Stresses in Heavy Closely Coiled Helical Springs", *Trans. ASME*, Vol. 51, Paper APM-51-17, 1929.
- 5 H. F. Ross, "Helical Spring Design Tables", *Machine Design*, p. #153, March, 1947.
- 6 H. F. Ross, "Application of Tables for Helical Compression and Extension Spring Design", *Trans. ASME*, Vol. 69, p. #725, 1947.
- 7 E. F. Smith, "Helical Compression Springs", *Machine Design*, April 19, 1956.
- 8 K. W. Maier, "Dynamic Loading of Compression Springs", *Product Engineering*, p. #162, March, 1955.
- 9 A. M. Wahl, *Mechanical Springs*, Penton Publishing Company, Cleveland, Ohio, 1949.

INDEX

Index Terms

Links

A

| | |
|---|--------|
| Alternating-current motors | 14-25 |
| comparison of typical motors | 14-37 |
| dynamic behavior of 2-phase servomotors | 14-31 |
| figure of merit | 14-37 |
| static characteristics of a-c motors | 14-31 |
| types of a-c motors used in servo- | |
| mechanisms | 14-25 |
| Alternative design methods | 6-15 |
| Amplifiers used in controllers | 13-1 |
| electronic amplifiers | 13-1 |
| hydraulic amplifiers | 13-105 |
| magnetic amplifiers | 13-58 |
| mechanical amplifiers | 13-181 |
| pneumatic amplifiers | 13-162 |
| relay amplifiers | 13-87 |
| rotary electric amplifiers | 13-73 |
| transistor amplifiers | 13-38 |
| Analog computers used for simulation | 3-39 |
| Analog-to-digital converters | 11-79 |
| coding discs | 11-81 |
| numerical representation | 11-79 |
| voltage-to-digital coders | 11-84 |
| Approximate numerical and graphical | |
| methods of determining transient response | 3-29 |
| Approximate procedures | 5-20 |
| approximate closed-loop response | 5-21 |
| general comments on the phase-margin | |
| criterion | 5-21 |
| phase margin and gain margin | 5-20 |

| <u>Index Terms</u> | <u>Links</u> |
|--|---------------------|
| Auxiliaries associated with servo mechanisms | 18-1 |
| auxiliary pumps | 18-1 |
| hydraulic auxiliaries | 18-4 |
| limit stops and positive stops | 18-15 |
| rotary joints | 18-12 |
| Auxiliary pumps | 18-1 |
| cost | 18-4 |
| leakage and drainage | 18-3 |
| maintenance | 18-3 |
| purpose | 18-1 |
| types of auxiliary pumps | 18-1 |
| B | |
| Bearings | 15-29 |
| ball bearings | 15-29 |
| friction | 15-32 |
| lubrication | 15-32 |
| miscellaneous bearings | 15-34 |
| roller bearings | 15-29 |
| sleeve bearings | 15-33 |
| Block diagrams and signal-flow graphs | 3-17 |
| block diagrams | 3-17 |
| signal-flow graphs | 3-25 |
| C | |
| Closed-loop response determination | 5-10 |
| gain-phase plane technique (Nichols chart) | 5-13 |
| nonunity-feedback systems | 5-15 |
| polar-plane technique | 5-11 |
| Compensation networks | 6-16 |
| a-c electric | 6-17 |
| d-c electric | 6-16 |
| hydraulic amplifier | 6-20 |
| mechanical damper | 6-19 |
| pneumatic controller | 6-21 |

Index Terms

Links

| | |
|--|-------|
| Compensation techniques | 6-1 |
| alternative design methods | 6-15 |
| feedback or parallel compensation | 6-10 |
| phase-margin and asymptotic methods | 6-6 |
| reshaping locus on gain-phase plane | 6-2 |
| typical compensation networks | 6-16 |
| Constants of fluids and tubing used in | |
| hydraulic systems | 20-15 |
| bulk modulus | 20-15 |
| compressibility | 20-15 |
| elastic properties of tubing | 20-15 |
| specific gravity | 20-15 |
| Constructional techniques | 19-1 |
| basic considerations | 19-1 |
| component layout | 19-1 |
| shock isolation | 19-10 |
| vibration isolation | 19-8 |
| Convolution integral | 3-11 |
| D | |
| Damping and friction | 20-4 |
| friction | 20-8 |
| viscosity | 20-4 |
| Describing function procedures | 10-1 |
| Design techniques | 9-9 |
| Differential equations | 3-1 |
| Digital-to-analog conversion | 12-16 |
| Direct-current motors | 14-1 |
| conversion factors and units | 14-23 |
| d-c torque motors | 14-20 |
| dynamic characteristics of d-c motors | 14-11 |
| measurement of d-c motor parameters | 14-17 |
| static characteristics of d-c motors | 14-7 |
| usage of d-c motors | 14-1 |
| Dynamic response | 2-1 |
| forced response | 2-8 |
| frequency response | 2-7 |

Index Terms

Links

Dynamic response (*Cont.*)

| | |
|--------------------|-----|
| linearization | 2-1 |
| stochastic inputs | 2-8 |
| transient response | 2-5 |

E

| | |
|---|-------|
| Electronic amplifiers | 13-1 |
| cascading amplifier stages | 13-15 |
| d-c power supply for electronic amplifiers | 13-35 |
| feedback amplifiers | 13-23 |
| linear analysis of single-stage voltage amplifiers | 13-7 |
| power amplifiers | 13-13 |
| problems encountered in use as servo components | 13-26 |
| thyatron amplifiers | 13-28 |
| vacuum tubes | 13-1 |

F

| | |
|--|------|
| Factoring and characteristic parameters of response modes | 3-2 |
| characteristic parameters of response modes | 3-4 |
| factoring | 3-2 |
| Feedback or parallel compensation | 6-10 |
| Forced response | 2-8 |
| Frequency response | 2-7 |

G

| | |
|---|------|
| Gain determination | 5-1 |
| approximate procedures | 5-20 |
| closed-loop response determination | 5-10 |
| exact and asymptotic-logarithmic representations | 5-4 |
| performance criteria and definitions | 5-1 |

Index Terms

Links

Gain determination (*Cont.*)

| | |
|--|------|
| polar-plane representation | 5-3 |
| root-locus method | 5-23 |
| setting the gain for a specified M_p | 5-15 |

Gain determination for a specified M_p

| | |
|-------------------------------|------|
| gain-phase plane construction | 5-18 |
| polar-plane construction | 5-16 |

Gear trains

| | |
|--|-------|
| definitions | 15-1 |
| design fundamentals | 15-5 |
| gear types | 15-3 |
| nonideal characteristics of gears and gear trains | 15-12 |

| | |
|---------|------|
| purpose | 15-1 |
|---------|------|

Gyroscopes

| | |
|--|-------|
| application factors | 11-71 |
| basic types of gyro units | 11-66 |
| description and basic theory | 11-61 |
| gyro unit performance | 11-64 |
| indication of the true vertical | 11-74 |
| inertial-guidance application of gyro units | 11-73 |
| single-axis integrating gyro unit | 11-69 |
| single-axis rate gyro unit | 11-68 |
| single-degree-of-freedom gyro units | 11-68 |
| two-degree-of-freedom gyro units | 11-66 |
| typical applications | 11-66 |

H

Hydraulic amplifiers

| | |
|---|--------|
| dynamic response of rotary pump and load | 13-149 |
| dynamic response of translational amplifier and load | 13-150 |
| hydraulic-circuit elements | 13-155 |
| illustrative example | 13-158 |
| interaction of load and valve | 13-137 |

Index Terms

Links

Hydraulic amplifiers (*Cont.*)

| | |
|------------------------------------|--------|
| problems encountered in use of | 13-150 |
| rotary hydraulic amplifiers | 13-136 |
| translational hydraulic amplifiers | 13-110 |

Hydraulic auxiliaries 18-4

| | |
|---|-------|
| accumulators | 18-8 |
| check valves | 18-4 |
| hydraulic systems incorporating auxiliaries | 18-4 |
| pressure-regulating valves | 18-6 |
| pressure-relief valves | 18-6 |
| unloading valves | 18-10 |

Hydraulic motors 14-37

| | |
|--|-------|
| approximate dynamic behavior of | |
| hydraulic transmission | 14-43 |
| dynamic behavior of hydraulic | |
| transmissions | 14-39 |
| parameter evaluation for rotary | |
| motor | 14-43 |
| problems encountered with hydraulic | |
| motors | 14-44 |
| static characteristics of piston-type rotary | |
| motors | 14-37 |
| static equations of translatory motor | |
| (moving piston) | 14-39 |

L

Laplace and Fourier transforms 3-12

| | |
|------------------------------------|------|
| frequency response | 3-17 |
| solution of differential equations | 3-13 |
| theorems | 3-12 |

Limit stops and positive stops 18-15

| | |
|-----------------|-------|
| characteristics | 18-15 |
| limit stops | 18-16 |
| positive stops | 18-18 |
| purpose | 18-15 |

Linearization 2-1

| <u>Index Terms</u> | <u>Links</u> |
|--|---------------------|
| Linear variable differential transformers | 11-47 |
| design characteristics | 11-48 |
| Linkages and levers | 15-17 |
| basic purpose | 15-17 |
| examples | 15-19 |
| nonideal characteristics | 15-19 |
| practical applications | 15-17 |
| M | |
| Magnetic amplifiers | 13-58 |
| analytical representation of | 13-64 |
| basic considerations | 13-58 |
| construction of | 13-70 |
| performance of | 13-66 |
| principles of operation | 13-59 |
| specifications and design | 13-72 |
| typical circuits | 13-62 |
| Magnetic-particle clutches | 14-47 |
| advantages | 14-47 |
| description | 14-47 |
| disadvantages | 14-47 |
| dynamic behavior | 14-51 |
| life expectancy | 14-51 |
| methods of use | 14-47 |
| static behavior | 14-50 |
| Mass moment of inertia | 20-1 |
| Mechanical amplifiers | 13-181 |
| basic types | 13-181 |
| dynamic behavior of | 13-187 |
| other mechanical amplifiers | 13-191 |
| problems encountered with | 13-195 |
| static characteristics of | 13-184 |
| Mechanical auxiliaries used in controllers | 15-1 |
| bearings | 15-29 |
| gear trains | 15-1 |

Index Terms

Links

Mechanical auxiliaries used in controllers (*Cont.*)

| | |
|-----------------------------|-------|
| link ages and levers | 15-17 |
| mechanical coupling devices | 15-22 |
| mechanical differentials | 15-16 |
| sheaves and tapes | 15-20 |

Mechanical coupling devices 15-22

| | |
|------------------|-------|
| couplings | 15-22 |
| keys and splines | 15-25 |

Mechanical differentials 15-16

| | |
|-----------------------|-------|
| differential linkages | 15-17 |
| geared differentials | 15-16 |
| purpose | 15-16 |

Methods of determining dynamic response

| | |
|---|------|
| of linear systems | 3-1 |
| approximate numerical and graphical | |
| methods of determining transient response | 3-29 |
| block diagrams and signal-flow graphs | 3-17 |
| convolution integral | 3-11 |
| differential equations | 3-1 |
| error coefficients for determining response | |
| to an arbitrary input | 3-34 |
| factoring and characteristic parameters of | |
| response modes | 3-2 |
| Laplace and Fourier transforms | 3-12 |
| response to stationary stochastic inputs | 3-36 |
| use of analog computers for simulation | 3-39 |

Modulators 12-2

| | |
|------------|------|
| chopper | 12-2 |
| electronic | 12-7 |
| magnetic | 12-4 |

N

Nonlinear systems 10-1

| | |
|--------------------------------------|-------|
| describing function procedures | 10-1 |
| limitations, compensation, and other | |
| methods | 10-13 |
| phase-plane procedures | 10-1 |

Index Terms

Links

O

| | |
|---|------|
| Objectives of feedback control system | 1-1 |
| Open-loop vs closed-loop system | |
| characteristics | 1-2 |
| Operational methods | 9-6 |
| additional properties of sampled | |
| functions | 9-8 |
| basic relations of sampled functions | 9-6 |
| Optimization methods for transient and | |
| stochastic inputs | 8-1 |
| criteria of performance | 8-1 |
| limitations and application problems | 8-11 |
| optimum synthesis of fixed-configuration | |
| systems | 8-2 |
| optimum synthesis of free-configuration | |
| systems with stationary stochastic inputs | 8-8 |
| Optimum synthesis of fixed-configuration | |
| systems | 8-2 |
| stationary stochastic inputs | 8-6 |
| transient inputs | 8-2 |
| Optimum synthesis of free-configuration | |
| systems with stationary | |
| stochastic inputs | 8-8 |

P

| | |
|--------------------------------------|-----|
| Performance criteria | 8-1 |
| Performance criteria and definitions | 5-1 |
| acceleration constant | 5-1 |
| bandwidth | 5-2 |
| gain | 5-1 |
| peak magnitude | 5-3 |
| static accuracy | 5-2 |
| torque constant | 5-2 |
| velocity constant | 5-1 |

| <u>Index Terms</u> | <u>Links</u> |
|--|---------------------|
| Performance evaluation | 7-1 |
| error coefficients | 7-44 |
| performance indices | 7-45 |
| relations between frequency response and transient response | 7-1 |
| Performance indices | 7-45 |
| Phase-margin and asymptotic methods | 6-6 |
| lag compensation | 6-7 |
| lead compensation | 6-7 |
| Phase-plane procedures | 10-8 |
| Pneumatic amplifiers | 13-162 |
| advantages and disadvantages of pneumatic systems | 13-180 |
| dynamic behavior of | 13-173 |
| pneumatic valves | 13-163 |
| static characteristics of pneumatic valves | 13-163 |
| Pneumatic motors | 14-44 |
| difficulties encountered with | 14-46 |
| dynamic behavior | 14-45 |
| principal types | 14-44 |
| static behavior | 14-45 |
| Polar-plane representation | 5-3 |
| direct polar plane | 5-3 |
| inverse polar plane | 5-4 |
| Potentiometers | 11-1 |
| application factors | 11-11 |
| description and basic theory | 11-1 |
| linear potentiometers | 11-4 |
| nonlinear potentiometers | 11-7 |
| Power elements used in controllers | 14-1 |
| alternating-current motors | 14-25 |
| direct-current motors | 14-1 |
| hydraulic motors | 14-37 |
| magnetic-particle clutches | 14-47 |
| pneumatic motors | 14-44 |

Index Terms

Links

| | |
|---|-----|
| Properties of feedback control systems | 1-1 |
| objectives of a feedback control system | 1-1 |
| open-loop vs closed-loop system | |
| characteristics | 1-2 |
| stability and dynamic response | 1-2 |
| terminology of feedback control systems | 1-3 |

R

| | |
|--|--------|
| Relations between frequency response and | |
| transient response | 7-1 |
| closed-loop frequency response from | |
| closed-loop transient response | 7-1 |
| relations between closed-loop transient | |
| response and closed-loop pole-zero | |
| configuration | 7-17 |
| relations between open-loop frequency | |
| response and closed-loop transient | |
| response | 7-21 |
| Relay amplifiers | 13-87 |
| advantages and disadvantages | 13-87 |
| dynamic characteristics of relays | 13-100 |
| parameter measurement | 13-101 |
| polarized relays | 13-93 |
| problems encountered with | 13-102 |
| relay characteristics | 13-88 |
| reversible motor-shaft rotation | 13-90 |
| single-sided relay amplifiers for speed | |
| control | 13-88 |
| static characteristics of relays | 13-98 |
| Representative designs | 17-1 |
| power control system for the M-38 fire- | |
| control system | 17-5 |
| servo system for a tracking-rauai | |
| antenna | 17-1 |
| Reshaping locus on gain-phase plane | 6-2 |
| lag compensation | 6-2 |
| lead compensation | 6-4 |

Index Terms**Links**

| | |
|---|-------|
| Response to stationary stochastic inputs | 3-36 |
| Root-locus methods | 5-23 |
| gain determination in the s plane | 5-24 |
| properties of roots in the s plane | 5-23 |
| Rotary electric amplifiers | 13-73 |
| characteristics of | 13-77 |
| parameters of d-c rotary amplifiers | 13-80 |
| problems encountered with the use of rotary electric amplifiers in servo applications | 13-83 |
| selection of rotary electric amplifiers for control purposes | 13-85 |
| types of | 13-73 |
| typical characteristics and design data of some rotary electric amplifiers | 13-86 |
| Rotary joints | 18-12 |
| dynamic seals | 18-12 |
| Rotary transformers | 11-15 |
| general classifications | 11-15 |
| general description | 11-15 |
| induction potentiometers | 11-41 |
| induction resolvers | 11-37 |
| microsyns | 11-44 |
| synchros | 11-16 |
| toroid-wound rotary transformers | 11-41 |
| Routh criterion | 4-2 |

S

| | |
|-------------------------------------|------|
| Sample-data systems | 9-1 |
| design techniques | 9-9 |
| general theory | 9-1 |
| operational methods | 9-6 |
| performance evaluation | 9-10 |
| z transform and the w transform | 9-4 |

| <u>Index Terms</u> | <u>Links</u> |
|---|---------------------|
| Sensing elements | 11-1 |
| analog-to-digital converters | 11-79 |
| gyroscopes | 11-60 |
| linear variable differential transformers | 11-47 |
| other forms of sensing elements | 11-86 |
| potentiometers | 11-1 |
| rotary transformers | 11-15 |
| tachometer generators | 11-51 |
| Sheaves and tapes | 15-20 |
| compliance | 15-21 |
| purpose | 15-20 |
| sheave sizes | 15-21 |
| stress | 15-20 |
| tension | 15-21 |
| Shock isolation | 19-10 |
| Signal converters | 12-1 |
| digital-to-analog conversion | 12-16 |
| electronic demodulators | 12-11 |
| modulators | 12-2 |
| types | 12-1 |
| Springs | 20-9 |
| cantilever springs | 20-11 |
| helical springs | 20-9 |
| torsion bar springs | 20-11 |
| vibration in springs | 20-14 |
| Stability and dynamic response | 1-2 |
| Stability of feedback control systems | 4-1 |
| Nyquist criterion | 4-4 |
| root-locus method | 4-7 |
| Routh criterion | 4-2 |
| Stochastic inputs | 2-8 |
| Supplementary tables, formulas, and | |
| charts | 20-1 |
| constants of fluids and tubing used in | |
| hydraulic systems | 20-15 |

Index Terms

Links

Supplementary tables, formulas, and (*Cont.*)

| | |
|------------------------|------|
| damping and friction | 20-4 |
| mass moment of inertia | 20-1 |
| springs | 20-9 |

T

| | |
|--|-------|
| Tachometer generators | 11-51 |
| d-c tachometer generators | 11-58 |
| drag-cup a-c tachometer generator | 11-51 |
| Terminology of feedback control systems | 1-3 |
| Transient response | 2-5 |
| Transistor amplifiers | 13-38 |
| a-c power amplifier design | 13-55 |
| analysis of transistor characteristics | 13-43 |
| basic principles | 13-38 |
| basic theory of junction diodes and | |
| transistors | 13-40 |
| transistor amplifier circuits | 13-50 |
| Typical procedure | 16-1 |
| analysis of trial system | 16-13 |
| choice of trial components | 16-4 |
| construction and test of experimental | |
| equipment | 16-14 |
| gathering specifications | 16-2 |
| illustrative example | 16-14 |
| modification or redesign of trial system | 16-13 |
| translation of experimental equipment | |
| into production model | 16-14 |

V

| | |
|---------------------|------|
| Vibration isolation | 19-8 |
|---------------------|------|

W

| | |
|---------------|-----|
| w transform | 9-5 |
|---------------|-----|

Z

| | |
|---------------|-----|
| z transform | 9-4 |
|---------------|-----|